Statistical Cross-Situational Learning to Build Word-to-World Mappings

Chen Yu (chenyu@indiana.edu)
Department of Psychological and Brain Sciences, Indiana University
Bloomington, IN 47405 USA

Linda B. Smith (smith4@indiana.edu)
Department of Psychological and Brain Sciences, Indiana University
Bloomington, IN 47405 USA

Abstract

There are an infinite number of possible word-to-world pairings in naturalistic learning environments. Previous proposals to solve this mapping problem focus on linguistic, social, representational constraints at a single moment. This paper investigates a cross-situational learning strategy based on computing distributional statistics across words, across referents, and most importantly across the co-occurrences of these two at multiple moments. We briefly exposed adults to a set of trials containing multiple spoken words and multiple pictures of individual objects with no information about word-picture correspondences within a trial. Nonetheless, subjects learned over trials the word-picture mappings through cross-trial statistical relations. Different learning conditions compared the degree of within-trial reference uncertainty, the number of trials and the length of trials. We also propose and implement a computational model and feed it with the same training data used in different learning conditions in experimental studies, to shed light on the possible underlying mechanism of statistical learning. Overall, these results suggest that statistical cross-situational learning may be one of fundamental mechanisms to tackle the word-to-world mapping problem.

Introduction

Children learn words in ambiguous contexts, with multiple word candidates for any referent and multiple referent candidates for any word. For example, a child may see a boy, a bat, a ball, and a dog and hear "Look at the boy. The dog wants his ball." This is the *word-to-world mapping* problem (e.g. Gleitman, 1990; Bloom, 2000; Smith, 2000). How could a learner who knows no words associate object names with the right referents? Developmentalists have studied a number of solutions to this problem, including ways in which the mature partner limits words and referents and directs attention to the relevant referent (Baldwin, 1993; Tomassalo, 2000), and internal perceptual and conceptual constraints (Genter, 1982). This paper is concerned with an additional solution, cross-situational statistical learning, a process in which statistics are calculated across different learning instances to determine *across* multiple experiences, the most likely word-referent mappings. We are also interested in how internal constraints, such as whole object assumption or mutual exclusivity, may be realized or embedded in these mechanisms.

Prior research has concentrated on *in-the-moment* solutions to the mapping problem. For example, the mutual exclusivity constraint (Markman, 1990) is hypothesized to direct children to map novel words to unnamed referents. If there are two objects present and one has a known name, the child should map a novel name to the second object, solving the word-referent mapping problem in that moment. Does this kind of constraint also contribute, perhaps in a graded way, over multiple encounters with words and potential referents? Children could use broader statistical regularities, keeping track of the associations among many words and referents across trials, using these, and adjusting these, as they encounter potential words and referents. The idea that the learning system may effectively calculate broad cross-situational statistics is suggested by recent findings on statistical learning in infants (Saffran, Aslin, & Newport, 1996). Infants (and children, adults, and nonhuman primates) readily learn transitional probabilities among segments in a temporal stream of syllables, tones, or visual events (Saffran, Johnson, Aslin, & Newport, 1999; Hauser, Newport, & Aslin, 2001; Newport & Aslin, 2004; Conway & Christiansen, 2005). All these studies concerned sequential statistics in streams of repeating segments. Here we examine a different kind of statistical learning – the mapping of units between a word and a referent stream.

Because this is the first investigation of this kind of statistical learning, we chose to study adult language learners, asking whether they could compute such statistics over many potential words and referents and asking the nature of the mechanisms that underlie such learning. We first present 2 experiments examining the capacities and limits of this learning. We then present a simulation study that explicitly examines how internal constraints such as the proposed mutual exclusivity assumption may be embedded in these statistical mechanisms.

Experiment 1

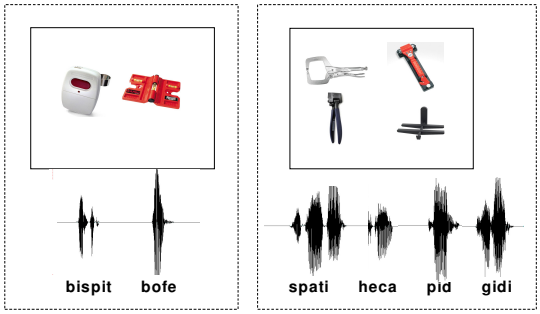
The power of statistical learning to overcome the mapping problem rests on the calculation of cross-situational statistics -- not just tracking, for example, the co-occurrences of ``ball" with ball or ``cup" with cup but the co-occurrences of ``ball" with a scene containing balls and dogs, balls alone, cups, cups and dogs, and so forth. Is this kind of computational mechanism at all feasible for humans? To answer this question, adult subjects were exposed to multiple trials wherein they heard multiple spoken words while looking at multiple pictures of objects. There is a perfect one-to-one mapping of words to referents such that each of the heard words maps to one of the objects. However, each trial consists of multiple words and multiples pictures of objects and there is no information

within a trial about the associations between words and referents (including no spatial or temporal cues). We manipulated the degree of ambiguity of each learning trial, presenting in one condition 4 words and 4 possible referents on each trial (16 potential associations), 3 words and 3 possible referents on each trial (9 potential associations), or 2 words and 2 possible referents (4 possible associations).

Method

Participants. 38 undergraduate and graduate students at Indiana University were tested in the experiment. Subjects received course credits or $7 for their participation.

Stimuli. Subjects were exposed to three learning conditions, each of which included 18 novel word-object pairs. In total, stimuli consisted of 54 visual-audio pairs in three conditions. The potential words were generated from a computer program to sample broadly from the space of phonotactically probable English. These artificial words were then produced by a synthetic female voice, presented in a monotone. 54 pictures of uncommon objects served as the visual input. The training trials were generated by pairing each word with a single picture. For each training trial, some number (depending on condition) of word-referent pairs were selected. Specifically, on each trial the referents were simultaneously presented on the screen. The names were then presented; however, the temporal order of the spoken names was not related in any systematic way to the spatial location of the referents. This is illustrated for a condition with 2 word-referent pairs and for a condition with 4 word-referent pairs in Figure 1. A 1000 ms silence was inserted between spoken words.



(a) 2 × 2 condition (b) 4 × 4 condition

Figure 1: Subjects saw multiple pictures while hearing multiple words on each trial, and were asked to find which spoken word is paired with which picture.

In total, there were three conditions determined by the number of words and referents presented on each trial: 2×2 (2 words and their corresponding referents), 3×3 (3 words and their corresponding referents), and 4×4 (4 words and their corresponding referents). In each condition, there were 18 unique word-picture pairs, and each unique word and corresponding unique referent were presented on a total of 6 training trials. This means, as shown in Table 1, that the total number of trials (of 2 pairs, 3 pairs or 4 pairs) is different over the three conditions. In order to keep the total training time (summed over all trials) constant, we also varied, as shown in Table 1, the length of time of each trial.

Table 1: three learning conditions in Experiment 1.

condition	# of total words	# of occ. per words	# of trials	time per trial (sec)	total time
2 × 2	18	6	54	6	324
3 × 3	18	6	36	9	324
4 × 4	18	6	27	12	324

Procedure. Visual stimuli were presented by 17 inch LCD flat panel screen and the sound was played by a pair of speakers connected to the same Windows PC. Subjects were instructed to map the pictures of objects showed on the computer screen onto the spoken words in a "nonsense" language. They were told that multiple words and pictures co-occurred on each individual trial and their task was to figure out which word went to which picture across multiple trials. Subjects were asked to participate in three sessions sequentially that corresponded to the three learning conditions. The order of sessions was counterbalanced. After each training session, subjects received a four-alternative forced-choice test consisting of 18 trials. For each testing question, subjects heard one word and were asked to select the corresponding picture from four options on the computer screen.

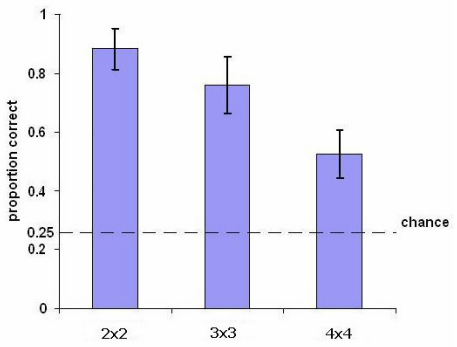


Figure 2: The results of three learning conditions in Experiment 1. Error bars reflect standard errors.

Results and Discussion

The three conditions present learners with different degrees of *in-trial* ambiguity but the same across trial (for a perfect statistical learner) certainty of word-referent pairings. The 4×4 condition –with four labels and four candidate referents for each learning moment (and thus 16 potential associations)– presents the greatest in-trial uncertainty, the 3×3 condition next, and the 2×2 condition least. As shown in Figure 2, in-trial uncertainty appears a relevant factor (ANOVA test, $F(2,74)=76.069$, $p<0.001$): Learners were better able to discover the correct word-referent in the 2×2 condition ($M=15.897$, $SD=2.506$) and least able in the 4×4 condition ($M=9.461$, $SD=2.907$) with performance in the 3×3 condition falling in between ($M=13.692$, $SD=3.507$). However, the most important result is that in all conditions, including 4×4, subjects performed reliably above chance ($t(37)=8.785$, $p<0.001$, one-tailed, for 4×4). Given the in-trial ambiguity, they must be calculating statistics across trials.

There are a number of potential explanations of the differences among the three conditions, including the central variability of degree of in-trial uncertainty but also the additional necessary confoundings of numbers of trials and length of trial. We investigate these factors in Experiment 2.

Experiment 2

Experiment 2 was designed to replicate the findings in Experiment 1 and further investigate under what circumstances, subjects would be able to achieve significantly better performance in the most ambiguous condition of Experiment 1, the 4×4 condition in which each trial offered 16 possible word-referent associations. In contrast to Experiment 1, and as summarized in Table 3, we probe the nature of statistical learning by manipulating two aspects of the training regime: : (1) the number of repetitions of each word-referent pair and (2) the total number of word-referent pairs to be learned.

Method

Participants. 28 undergraduate students at Indiana University were tested in this experiment. None of them participated in Experiment 1.

Stimuli. all three conditions in Experiment 2 use the 4×4 presentation of 4 words and 4 pictures on each trial. However, in the 9 words, 8 repetitions condition, subjects attempt to discover a total of 9 word-referent pairs each repeated 8 times over the course of training. In the 9 words, 12 repetitions condition, subjects attempt to discover 9 word-referent pairs but are given 4 additional repetitions of each word-referent pair. Finally, the third condition is a replication of the 4×4 condition of Experiment 1; there are 18 word-referent pairs to be learned and 6 repetitions of each. In contrast to Experiment 1, total viewing time per slide was kept constant and thus the total number of trials and total length of the experiment varied across conditions.

Table 2 The statistics of the stimuli in 3 learning conditions.

learning condition	# of total words	# of occ. per word	# of trial	time per trial	total time
9 words/ 8 repetitions	9	8	18	12	216
9 words/ 12 repetitions	9	12	27	12	324
18 words/ 6 repetitions	18	6	27	12	324

Procedure. The procedure is the same with that of Exp. 1.

Results and Discussion

There were 4 words and 4 pictures in a single trial in the three conditions, which contained a high degree of ambiguity at each individual moment (trial). However, subjects in all conditions again discover more pairs than expected by chance as shown in Figure 3 ($t(27) > 6.4$ in all three conditions). In addition, the results in 18 words/6repetions condition of this experiment (M=9.629, SD=3.076) are very similar to the same condition in Experiment 1, suggesting that our results are reliable and duplicable. The above two observations are quite in line with what we expected.

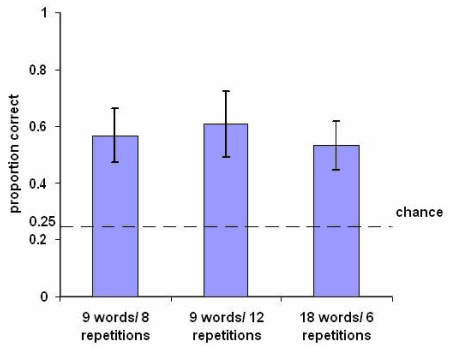


Figure 3: the results in three conditions in Experiment 2. Error bars reflect standard errors.

However, as shown in Figure 3, in all three conditions, subjects –in terms of the proportion of word-referent pairs to be discovered – performed equivalently ($F(2,54) = 0.052$; $p<0.001$). A direct comparison between 18 words/6 repetions (M=0.534, SD=0.232) and 9 words/12 repetitions (M=0.609, SD=0.176) conditions shows that they have the same number of trials, the same training time and the same within-trial ambiguity. The only two different factors are: (1) the number of unique pairs and (2) the number of occurrences per pair, which was expected to make one condition easier than the other. However, the results in these two conditions are quite similar. An intuitive explanation is that the number of co-occurring pairs plays a dominant role in statistical word learning and other factors are not so important conditioned on that factor. But why is that? Our following computational study provides a plausible answer to these behavioural data.

Simulation

The above two experiments document the performances of adults in statistical word learning. We demonstrated *what* they can do given cross-situational observations. The next question to ask is *how* they do that -- the underlying computational mechanisms that support statistical word learning. Since there is no information at the beginnings to guide them to discover correct word-referent pairs among all possibilities, they must start with randomly selecting some hypothesized pairs and then gradually justify the correctness of those pairs later. Following this general principle, the specific questions in statistical word learning are (1) how hypothesized pairs are selected and stored from a trial? (2) how subjects justify whether a word-object pair is correct? (3) whether they use the mutual exclusivity constraint if two working hypothesized pairs are not compatible? and (4) whether they would use previously learned pairs to help the learning of new pairs in subsequent trials? The following simulation study attempts to answer those questions by showing a dynamic picture of the real-time learning when the simulated learner is fed with the same stimuli that subjects were exposed.

Training:

-- Randomly select one pair from Trial #1 and store it in the memory as the first hypothesized pairing.

-- Repeat the following steps for Trial #i ($2 \le i \le 27$):

a. Check the pairs in the memory M and use those with a high confidence score $c_{n_j m_j}$ to filter the input of the current trial T_i.

b. Randomly selection a new pair (p_{new}, w_{new}) from T_i.

c. Comparing (p_{new}, w_{new}) with the parings in M:

if $(p_{new}, w_{new}) \in M$

Increase the confidence score of the corresponding pairing $c_{n_j m_j}$.

else if $p_{new} \notin \{p_{n_1}, p_{n_2}, ..., p_{n_k}\}$ and $w_{new} \notin \{w_{n_1}, w_{n_2}, ..., w_{n_k}\}$

Add the pair into M as $(p_{n_{k+1}}, w_{m_{k+1}}, c_{n_{k+1} m_{k+1}})$ and $c_{n_{k+1} m_{k+1}} = 1$.

else if $p_{new} \notin \{p_{n_1}, p_{n_2}, ..., p_{n_k}\}$ and $w_{new} \in \{w_{n_1}, w_{n_2}, ..., w_{n_k}\}$

Finding the pairing (p_{n_j}, w_{m_j}) in M while $w_{m_j} = w_{new}$.

If $p_{n_j} \in \{p_{i_1}, p_{i_2}, p_{i_3}, p_{i_4}\}$ then increase $c_{n_j m_j}$ by 1,

Otherwise replace $(p_{n_j}, w_{m_j}, c_{n_j m_j})$ with $(p_{new}, w_{new}, 1)$.

else if: $p_{new} \in \{p_{n_1}, p_{n_2}, ..., p_{n_k}\}$ and $w_{new} \notin \{w_{n_1}, w_{n_2}, ..., w_{n_k}\}$

Finding the pairing (p_{n_j}, w_{m_j}) in M while $p_{n_j} = p_{new}$.

If $w_{m_j} \in \{w_{i_1}, w_{i_2}, w_{i_3}, w_{i_4}\}$ then increase $c_{n_j m_j}$ by 1,

Otherwise replace $(p_{n_j}, w_{m_j}, c_{n_j m_j})$ with $(p_{new}, w_{new}, 1)$.

Testing:

For ith question $\{w_{i_1}, p_{i_1}, p_{i_2}, p_{i_3}, p_{i_4}\}$,

If $w_{i_1} \in \{w_{n_1}, w_{n_2}, ..., w_{n_k}\}$, find the corresponding pair (p_{n_j}, w_{m_j}) in M while $w_{m_j} = w_{i1}$ and check whether $p_{n_j} == p_{i_1}$;

Otherwise, among $\{p_{i_1}, p_{i_2}, p_{i_3}, p_{i_4}\}$, remove those $p_{i_{1-4}} \in \{p_{n_1}, p_{n_2}, ..., p_{n_k}\}$ and randomly select an answer from the left items.

Figure 4: Statistical cross-situational learning algorithm.

Method

The 4×4 condition has been tested in both experiments and therefore is used as an example to show how the model works. The simulations on other conditions can be achieved by applying the corresponding stimuli to the same model. In the 4×4 condition, the 18 novel word-picture pairs can be represented as $\{(p_1, w_1), (p_2, w_2), ...(p_{18}, w_{18})\}$. In the ith trial, the stimuli are $T_i = \{p_{i_1}, p_{i_2}, p_{i_3}, p_{i_4}, w_{i_1}, w_{i_2}, w_{i_3}, w_{i_4}\}$ while i_1, i_2, i_3 and i_4 can be selected from 1 to 18. And there is no information as to which picture goes with which name. We also assume that the simulated learner maintains a list of hypothesized pairings as learned results from previous trials. Moreover, the learner assigns a confidence score for each pair in his memory to indicate the likelihood that the pair is correct. His lexical knowledge at the ith trial can be then represented as a list of pairs $M = \{(p_{n_1}, w_{m_1}, c_{n_1 m_1}), (p_{n_2}, w_{m_2}, c_{n_2 m_2}), ..., (p_{n_k}, w_{m_k}, c_{n_k m_k})\}$ while n_j and m_j can be selected separately from 1 to 18, and $c_{n_j m_j}$ is the confidence score of a pair. Thus, the equivalence of n_j and m_j indicates a correct pairing.

At the beginnings, the model randomly picks one word and one picture from a trial and builds a hypothesized pairing. With more trials, more pairings are built and stored in the memory. Two additional mechanisms are utilized to make this learning process more effective. First, one important constraint in adding new pairs is to maintain the consistency of hypothesized pairings so that one word can be associated with only one picture. This constraint explicitly encodes the proposals such as mutual exclusivity (Markman, 1990) and contrast (Clark, 1987) into the learning machinery and by doing so makes the learning more efficient because the simulated learner would randomly select many conflicting (and therefore incorrect) word-picture pairs across multiple trials without this constraint. Second, the model keeps track of the confidence score of each pair. When the confidence score of a pair is above a certain threshold, this pair will be treated as a learned lexicon and then used to filter out the input in subsequent trials, which can significantly simplify the learning task. For instance, if a learned pair occurs in a new trial, it will be removed from the stimuli to reduce a 4-4 condition into a 3-3 condition. More importantly, subjects in empirical studies informed experimenters that they used the similar filtering strategy in the later part of the training phase when they were confident that some word-picture pairs were correct. The detailed learning algorithm is described in Figure 4.

Results and Discussion

We applied the same training and testing data in the previous experiments to the simulated learner. For each condition, the simulation was run for 5000 times. Thus, we had 5000 simulated subjects (with the same set of parameters) for each condition. Note that the fundamental mechanism encoded in our model is to randomly select and store hypothesized pairs. Therefore, quite different results were obtained on each run depending on what pairs were selected from trial to trial. We used 5000 simulated subjects

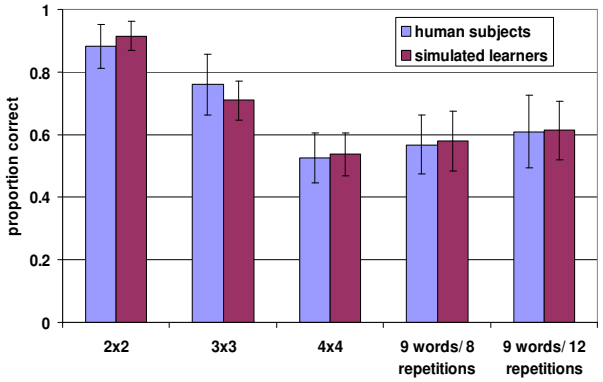


Fig 5: A comparison of human subjects and simulated learners.

to ensure the statistical power of this simulation study and the results are shown in Figure 5.

We observed that in general the results in simulation are quite in line with those of human subjects, suggesting that if subjects apply a simple statistical learning machinery like the one in our model, then that could explain their superior performances. Admittedly, different subjects may apply different learning strategies and there is no way to encode all the possible strategies that they applied to the stimuli in a single model. Nonetheless, the current model intends to implement one learning device based on general principles. The similarities of the results between human subjects and simulated subjects are consistent not only in one condition but among multiple conditions, indicating that the learning principles encoded in our model are plausible to be similar with those guiding the learning of human subjects. We will discuss why simulated learners couldn't achieve a better performance in the 9 pairs/12 repetitions condition in the next section.

We also note that one reason for individual differences in this type of learning task is that if learners just randomly select pairs and justify them later based on distributional information, then the results obtained from different trials of running the same model could be quite different. In some cases, simulated learners may happen to pick correct pairs in the first trials which will help subsequent learning through the filtering mechanism. In other cases, they might pick the wrong ones and have to justify and correct those pairs in the later trials. Thus, the randomness in pair selection may also cause those human subjects, who apply the same learning mechanism on the same data, to achieve quite different results.

General Discussion

Statistical Learning

The learning situations such as those used in the present experiments have generally been considered too complex for word learning. Yet the present results show that adults rapidly discover word-referent mappings in these contexts. The only solution to the mapping problem is the distributional co-occurrence statistics between spoken words and pictures of objects. Our findings in statistical word learning extend those of Saffran, Aslin, & Newport (1996), and Newport and Aslin (2004) in word segmentation, Gomez & Gerken (1999) in syntax learning, and Conway & Christiansen, (2005) in visual and tactile sequence learning by showing that statistical learning broadly characterizes human learning, and that human learners can exploit cross-trial regularities over many potential word and referent pairs.

Conclusions relevant to development are limited by our use of adult subjects. Nonetheless, recent studies in word learning (e.g. Gillette, Gleitman, Gleitman, & Lederer, 1999; Snedeker & Gleitman, 2004) proposed a Human Simulation Paradigm (HSP), suggesting the value of examining potential general learning mechanisms in adults as a way to study the potency of various cues to word learning that might be available in the learning environment. Therefore, adult studies can be used as first steps and proof-of-concept before conducting infant studies (Saffran, Newport, & Aslin, 1996; Newport & Aslin, 2004). In fact, our on-going studies on young children apply the same experimental paradigm and the same visual-auditory stimuli used in the present study.

Key Factors in Cross-Situational Learning

The five learning conditions in two experimental studies provided different kinds of statistical regularities for learning. We will focus on two important factors discovered through these manipulations.

First, the dominating factor in the cross-situational learning is the number of co-occurring word-picture pairs, a result confirmed in both experiments and simulation. This factor determines the probability of selecting a correct pair from a trial. Specifically, the probability of picking a correct pair is 2 out of 4 in the 2×2 condition, 3 out of 9 in the 3×3 condition, and 4 out 16 in the 4×4 conditions. Note that if subjects select an irrelevant pair, there are two possible consequences: (1) they select the corresponding correct pair later and based on the mutual exclusivity constraint, exclude the wrong pair; (2) they never receive evidence to justify the pairing so based on their limited exposure, they could either believe that the pairing is correct or ignore this pair in testing. To sum up, whenever subjects select a wrong pair, which is almost unavoidable, the pairing would either require their justification in subsequent trials to correct it in the best case or lead to wrong answers in testing in a worse case. Therefore, the probability of selecting a correct pair plays a key role in learning, no matter what learning algorithms are applied to those hypothesized pairs later.

The second factor is the total number of unique pairs. We found that the 9 pairs/12 repetitions condition is not significantly better than the 18 pairs/6 repetitions condition. One plausible reason is that the probabilities of selecting a correct pair from a trial in these two conditions are the same. With this low probability (=0.25), word learners would be likely to randomly select a wrong pair from a trial and later have to exclude it. From our simulation, we also found that with 4-pair in a trial and 9 word-picture pairs in total, it is more likely that two word-picture pairs (e.g. p_1-w_1 and p_2-w_2) co-occur more frequently across multiple trials in the 9 pairs/12 repetitions condition compared with selecting 4 out of 18 in the 18 pairs/6 repetitions condition. If word learners happen to select a wrong pairing (e.g. p_1-w_2 or p_2-w_1) in multiple times from those trials, then they may "confidently" reach wrong conclusions. Thus, the fewer number of word-picture pairs in total causes irrelevant (false) word-referent pairs to repeatedly co-occur in multiple trials, which may mislead word learners if they happen to pay attention to wrong pairings. This in fact makes the learning situation harder but not easier. The claim that more pairs are better than fewer in statistical learning sounds quite controversial. This is somehow an intriguing and compelling finding.

Modeling Statistical Cross-Situational Learning

Our model encodes a very fundamental (and rather simple) mechanism – building one hypothesized pair from a trial, saving it in the memory, gradually adding more pairs, justifying the correctness of a hypothesized pair in subsequent trials, and removing conflicting pairs if needed. However, similar to adult learners, the model achieved quite impressive performances in highly ambiguous learning conditions, suggesting that a powerful statistical learning capability can be achieved by a relatively simple learning mechanism. We also argue that this general learning principle has been applied more or less by human subjects. They may differ in the number of hypothesized pairs they could select and memorize from a trial, or in the way to decide which pairs to be selected, or in how many hypothesized pairs could be saved in the memory, or in when and how to justify those hypothesized pairs in the memory. Nonetheless, the general learning mechanism could be quite similar to the model described above and all those factors mentioned above can be treated as the parameters of this general learning model. Two important mechanisms that make the model work more effectively are (1) one-word-to-one-object constraint and (2) using learned pairs to filter new input. We noticed that the one-to-one constraint is especially useful in learning. Compared with previous studies, we demonstrate, both experimentally and computationally, the role of this constraint across multiple trials (but not in a single moment), the learning situation that is more representative of naturalistic learning environments that children are situated in. The fact that the mechanism encoding with this type of constraint achieved similar learning performance with human subjects indicates the plausibility that subjects utilize at least similar (if not identical) constraints. Moreover, by feeding the model with the same stimuli that subjects were exposed in the training phases and asking the model to do the same tests after training, we can demonstrate moment-to-moment changes in statistical learning processes that human subjects might experience.

Conclusion

Previous studies show that statistical learning is applied in various learning tasks, such as speech segmentation, syntactic learning and visual processing. In light of this, we study to what degree language learners can acquire word-to-world mappings through statistical regularities in co-occurring visual-auditory streams. We showed that a significant amount of lexical knowledge can be learned through statistical learning. Moreover, we systematically manipulated different statistical properties of the stimuli and measure the learning capacities of adult learners in various learning conditions. To understand underlying learning mechanisms, we developed a computational model that was fed with the same stimuli of human subjects, and simulated learners achieved similar performances as human learners. In this way, we obtained a more complete picture of statistical word learning. Our next step is to extend current studies to infants and young children to investigate how well they could utilize statistical cues to tackle the word-to-world mapping problem.

References

Baldwin, D. (1993). Early referential understanding: Infant's ability to recognize referential acts for what they are. *Developmental psychology* (29), 832-843.

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

Conway, C. & Christiansen, M.H. (2005). Modality constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory & Cognition, 31,* 24-39.

Clark, E.V. (1987). The Principle of Contrast: a constraint on language acquisition. In B. MacWinney (Ed.), *Mechanisms of language acquisition* (pp. 1-33): Hillsdale, NJ: Lawrence Erlbaum Associates.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj II (Ed.), *Language development* (Vol. 2). Hillsdale, NJ: Erlbaum.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition, 73*, 135-176.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition, 1* 1-55.

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*(2), 109-135.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition, 78*(3), B53-64.

Markman, E. M. (1990). Constraints Children Place on Word Learning. *Cognitive Science, 14*, 57-77.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*(2), 127-162.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27-52.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word Segmentation: The role of distributional cues. *Journal of memory and language, 35*, 606-621.

Smith, L. B. (2000). How to learn words: An Associative Crane. In R. Golinkoff & K. Hirsh-Pasek (Eds.), *Breaking the word learning barrier* (pp. 51-80): Oxford: Oxford University Press.

Snedeker, J., & Gleitman, L. (2004). Why it is hard to label our concepts. In G. Hall & S. R. Waxman (Eds.), *Weaving a Lexicon*. Cambridge,MA: MIT Press.

Tomasello, M. (2000). Perceiving intentions and learning words in the second year of life. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 111-128): Cambridge University Press.