CHAPTER TWO

Mechanisms of Cross-situational Learning: Behavioral and Computational Evidence

Yayun Zhang[a], **Chi-hsin Chen**[b], **Chen Yu**[a,*]
[a]Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, United States
[b]Department of Otolaryngology-Head and Neck Surgery, The Ohio State University, Columbus, OH, United States
*Corresponding author: e-mail address: chenyu@indiana.edu

Contents

Abstract

Word learning happens in everyday contexts with many words and many potential referents for those words in view at the same time. It is challenging for young learners to find the correct referent upon hearing an unknown word at the moment. This problem of referential uncertainty has been deemed as the crux of early word learning (Quine, 1960). Recent empirical and computational studies have found support for a statistical solution to the problem termed cross-situational learning. Cross-situational learning allows learners to acquire word meanings across multiple exposures, despite each individual exposure is referentially uncertain. Recent empirical research shows that infants, children and adults rely on cross-situational learning to learn new words (Smith & Yu, 2008; Suanda, Mugwanya, & Namy, 2014; Yu & Smith, 2007). However, researchers have found evidence supporting two very different theoretical accounts

Advances in Child Development and Behavior, Volume 56
ISSN 0065-2407
https://doi.org/10.1016/bs.acdb.2019.01.001

of learning mechanisms: Hypothesis Testing (Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005; Markman, 1992) and Associative Learning (Frank, Goodman, & Tenenbaum, 2009; Yu & Smith, 2007). Hypothesis Testing is generally characterized as a form of learning in which a coherent hypothesis regarding a specific word–object mapping is formed often in conceptually constrained ways. The hypothesis will then be either accepted or rejected with additional evidence. However, proponents of the Associative Learning framework often characterize learning as aggregating information over time through implicit associative mechanisms. A learner acquires the meaning of a word when the association between the word and the referent becomes relatively strong. In this chapter, we consider these two psychological theories in the context of cross-situational word-referent learning. By reviewing recent empirical and cognitive modeling studies, our goal is to deepen our understanding of the underlying word learning mechanisms by examining and comparing the two theoretical learning accounts.

1. The problem: Word learning challenge

Young children are skilled word learners. During the second year of life, the rate at which children acquire new words accelerates dramatically (McMurray, 2007). By age of 18, they know around 60,000 words (Bloom, 2000) including nouns, verbs, adjectives, and other word classes. However, the set of words acquired before age 2 contains a large proportion of common nouns for concrete things (Fenson et al., 1994; Gentner, 1982; Markman, 1989). In order to acquire the meanings of their first words, young learners have to rely on observing the immediate context to identify relevant information upon hearing a word. However, determining the meaning of a newly encountered noun is challenging because in principle there is an unlimited number of referents inherent in the learning moment (Quine, 1960). How do infants know which word label maps onto which object? This problem is termed "referential uncertainty" in the word learning literature and has been studied extensively in the past several decades.

Carey and Bartlett's (1978) seminal study of fast mapping suggests that children as young as 3 years of age can identify at least some aspects of the meaning of a novel word after only a few exposures, and they demonstrated successful retention of the novel word 1 week later. Numerous studies over the past several decades have replicated the general finding of fast mapping, and many have focused on how referential uncertainty can be reduced precisely when parents name an object. For example, children use behavioral cues to identify the speaker's referential intention (Baldwin, 1991; Tomasello & Farrar, 1986); they assume a word refers to the whole object rather than parts or properties of the object (MacNamara, 1972); and they assume words have mutually exclusive meanings, therefore existing knowledge can be helpful for finding new word meanings (Markman & Wachtel, 1988). In addition, argument structure and syntactic context could also facilitate word learning (e.g., Gillette, Gleitman, Gleitman, & Lederer, 1999). Young learners clearly are able to infer correct word-referent mappings in those referentially clear moments, but everyday learning contexts can be messy and highly ambiguous (Medina, Snedeker, Trueswell, & Gleitman, 2011). One critical question is whether young learners acquire word-referent mappings from ambiguous contexts, and if so, how.

Cross-situational learning (CSL) is a mechanism proposed to explain word learning in noisy contexts as learners aggregate information across individual learning situations (Siskind, 1996; Smith, Smith, & Blythe, 2011; Yu & Smith, 2007). The logic of CSL is that when learners hear a word, there is a set of potential candidate referents available. Although learners are unable to identify the correct word–object mapping after a single exposure, if learners can combine information across multiple exposures, they are able to determine the most probable referent by integrating multiple candidate sets over time. Basically, hearing words in enough various contexts would allow learners to rule out incorrect associates and learn the most consistent mappings, which are likely to be the correct ones. A growing body of research shows that adults are quite good at accumulating statistical evidence across individually ambiguous learning contexts with multiple novel words and multiple novel objects (Aussems & Vogt, 2015; Chen, Zhang, & Yu, 2018; Fitneva & Christiansen, 2011; Koehne & Crocker, 2015; Monaghan, Mattock, Davies, & Smith, 2015; Onnis, Edelman, & Waterfall, 2011; Wang & Mintz, 2018; Yu & Smith, 2007). Experimental studies also indicate that infants and young children do this kind of learning as well (Akhtar & Montague, 1999; Scott & Fisher, 2011; Smith & Yu, 2008; Vlach & Johnson, 2013; Vouloumanos & Werker, 2009; Suanda, Mugwanya, & Namy, 2014). To illustrate how CSL works, we use a simple example (Fig. 1): a learner hears the words "ball" and "bat" in the context of seeing object BALL and object BAT; without other information, the learner cannot know whether the word form "ball" refers to one or the other visual object. However, if subsequently, while viewing another scene with the potential referents of BALL and DOG, the learner hears the words "ball" and "dog" and if the learner can combine cooccurrence information from these two trials, the learner could correctly map "ball" to the object BALL (and perhaps also infer the connection between the word "bat" and the object BAT).

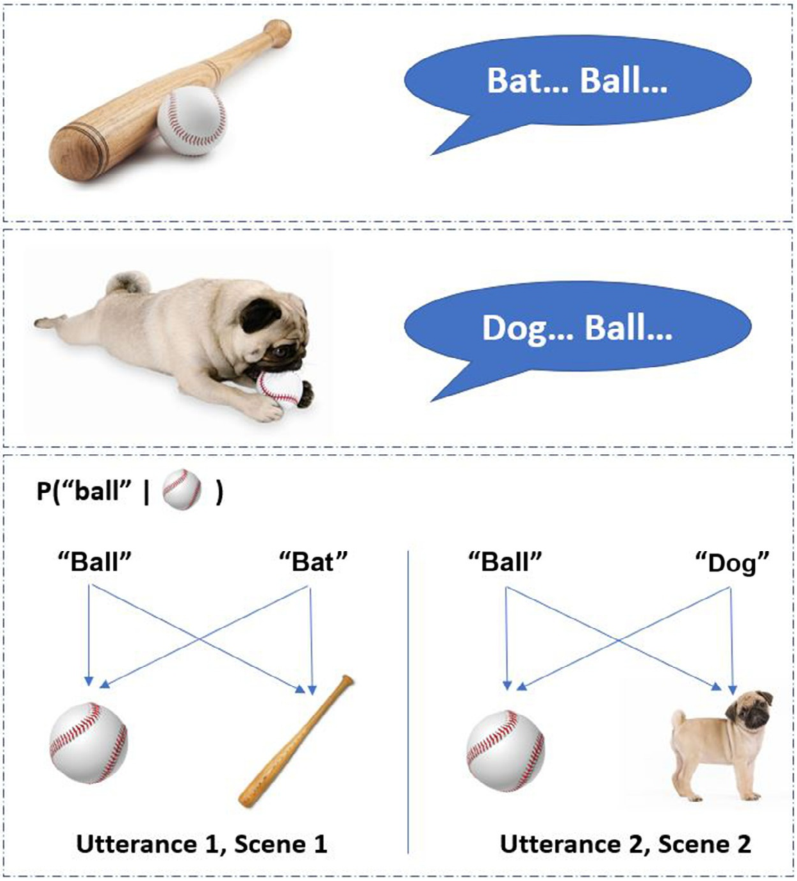


Fig. 1 A toy example of cross-situational learning to illustrate how words are learned with multiple learning trials.

Although the cross-situational solution seems to be quite straightforward, cross-situational statistical learners need to possess the following set of cognitive skills to accomplish the learning task: (1) they need to recognize both individual word tokens and individual objects; (2) they need to register cooccurrences and non-cooccurrences; (3) they need to remember previous word-referent pairings; (4) they need to aggregate the information across trials; and (5) finally, they need to calculate the correct statistics from cross-trial information. What is the mechanism or system of mechanisms that supports word-referent statistical learning? Extant computational and experimental research suggests that more than one mechanism could explain cross-situational word learning findings, including Hypothesis Testing (HT)

(Medina et al., 2011; Trueswell, Medina, Hafri, & Gleitman, 2013) and Associative Learning (AL) (Kachergis, Yu, & Shiffrin, 2012; Yu & Smith, 2007). The goal of this chapter is to better understand how different learning mechanisms work, how they account for empirical evidence collected from both infant and adult learners using different experimental paradigms, and to discuss possible new directions to advance our understanding of not only CSL but also early word learning in general.

2. Hypothesis testing vs associative learning

2.1 Hypothesis testing

One possible mechanism underlying CSL is called Hypothesis Testing (HT, Bloom, 2000; Carey & Bartlett, 1978; Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005; Markman, 1992). According to this account, when children encounter a new word, they make a specific hypothesis about what that word means. As they encounter the same word in a subsequent naming event, they either confirm the hypothesis (i.e., the same object is present in the subsequent event) or reject the hypothesis (i.e., the same object is absent). In rejected cases, a new hypothesis will be proposed and tested in subsequent events (see Medina et al., 2011; Trueswell et al., 2013). As shown in the toy example in Fig. 2, the learners hear the word "ball" and "bat" while seeing both objects, then they make a hypothesis about the correct mapping by randomly associating an object (BALL) with a word (bat). In the next situation, they hear the word "ball" and "dog" while both objects are present, learners check to see whether the initial BALL—"bat" hypothesis can be confirmed. In this case, the word "bat" is absent, indicating that the initial mapping is wrong. Therefore, learners discard the BALL—"bat" hypothesis and randomly pick another

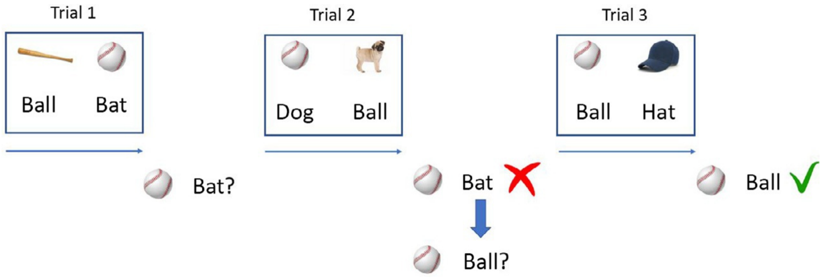


Fig. 2 A toy example of hypothesis testing.

one based on the current information. In this example, they pick BALL—"ball" as the new to-be-confirmed hypothesis. They then repeat the checking process in the third learning trial wherein the word "ball" and the object BALL are both present, confirming the second hypothesis, and thereafter the word is considered learned. One key feature of the HT account is that learners only form one hypothesis or conjecture about an object-label pairing in one learning moment. They only retain this hypothesis and discard all possible alternatives with no memory of previous experience. In extreme cases, learners could pick the same wrong mapping repeatedly. This HT model suggests that children begin word learning by making an initial fast mapping between a new word and its likely meaning. Although they also modify the guesses as more input comes in, they do not retain or use accumulated knowledge from past experiences.

2.1.1 Empirical evidence on HT

Most empirical evidence on HT was based on the "Human Simulation Paradigm (HSP)" pioneered by Gillette et al. (1999). In one study using HSP, researchers recorded videos of natural interactions between parents and their 12- to 15-month-old toddlers at home (Medina et al., 2011). The videos were cut into 40-s vignettes of parents uttering labels to their children. The videos were muted, and a beep sound was inserted when the parent labeled an object. Adult participants were asked to watch the videos and guess which object the parent intended to label at the beep. They found that 90% of the vignettes depicting natural learning instances had accuracy scores below 33%, which suggested that they were uninformative for identifying the correct referent of the label. Only a small percentage (7%) of naming instances were considered highly informative. In a following experiment, researchers showed participants five vignettes, in which four were Low Informative (LI) and one was High Informative (HI). Across four different conditions, the HI vignette was placed in different positions: at the beginning, in the middle, at last, and no HI. Participants made one guess of the target referent after viewing each vignette, and they also provided a final conjecture at the end of the experiment. The results showed that participants were only able to learn the correct word-referent mappings across trials in the condition where the HI trials were presented first. Learners reached high accuracy on the first HI vignette (66%) and maintained that guess by the fifth vignettes (41%). Performance was significantly worse when the first vignette was LI (below 20%) and learners failed to identify the word at the end even in cases that HI vignette was placed in the middle or last. Their findings suggest that learners only remembered their previous successful guess. If that guess was rejected, participants had little to no memory of alternative pairings that they could return to, and therefore could not improve their performance in later guesses. Guesses on the first trial appeared to determine later learning accuracy, which is consistent with the HT model (Medina et al., 2011).

2.1.2 Modeling work on HT

Researchers have also proposed and implemented several computational models built upon the HT approach. To illustrate, we use an extended version of the learning scenario before we discuss specific models. Let's consider a word-learning task as shown in Fig. 3. Learners hear the to-be-learned word "bosa" five times in different scenarios and each time they see three different objects. How do learners figure out which object is "bosa"?

According to an HT model called Propose-but-Verify (Trueswell et al., 2013), the learners would randomly pick an object as the referent first, in this case, they pick BALL. Because BALL is also present in the second trial, they were able to confirm this hypothesis and keep it in their working memory. Same confirmation can be made in the third trial. At trial 4, because BALL is no longer present, learners reject the original hypothesis and pick a new object from the current set (i.e., BAT). Now "bosa"—BAT is the new hypothesis and will be either rejected or confirmed in the subsequent trial depending on whether it is present. In general, this Propose-but-Verify algorithm follows three simple steps: (1) begin by randomly selecting a hypothesis; (2) when the word occurs again, remember the previous guess with some probability; and (3) if the selected pair is confirmed in the current referent set, the model would select the referent; otherwise discard the pair and select another referent at random. Trueswell et al. (2013) found that

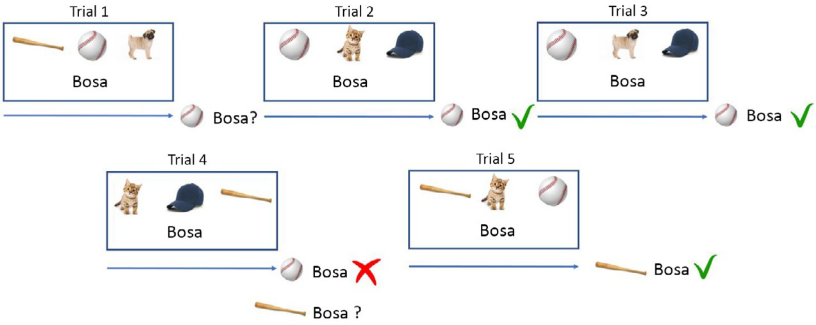


Fig. 3 Propose-but-Verify model.

the very simple Propose-but-Verify model with only one free parameter captured the general learning patterns observed in human data, providing support for the HT account.

A variant of the Proposed-but-Verify model called Pursuit is described in Stevens, Gleitman, Trueswell, and Yang (2017). Instead of completely getting rid of a disconfirmed hypothesis, the Pursuit model lowers the association value of disconfirmed pairs, which means they may still remain a probable hypothesis and could be selected in future trials (Stevens et al., 2017). As shown in Fig. 4, participants first select BALL—"bosa" as the initial hypothesis, and this hypothesis gets confirmed once in the second trial and again in the third trial. At trial 4, the BALL—"bosa" hypothesis is rejected because the object BALL is not present. Instead of completely rejecting this hypothesis, learners may instead lower its likelihood as the correct mapping and at the same time select another object (i.e., BAT) as a new possible referent. They store both hypotheses in memory and on subsequent trial 5, both hypotheses will be checked. In this case, BALL and BAT are both present; therefore, the association value for both hypotheses will be strengthened. Like the Propose-but-Verify model, the Pursuit model considers only a few hypotheses and ignores all other possibilities upon confirmation. However, unlike the Propose-but-Verify model, a disconfirmed referent is not discarded but only has its association value lowered, allowing the possibility to be selected if it remains the most probable hypothesis next time the word is presented. Both models suggest that additional referents are added to the hypothesis set only if the most favored referent fails to be confirmed. In other words, when the most favored referent continues to be confirmed, the learner ignores all other competing referents, even when they are also present. This is one of the key differences between Pursuit learning and AL, which will be discussed next.

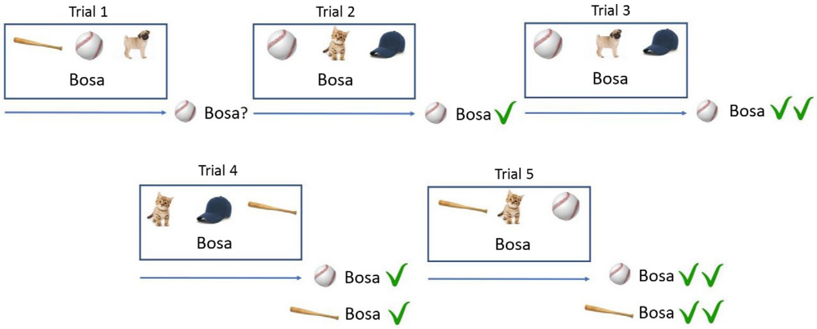
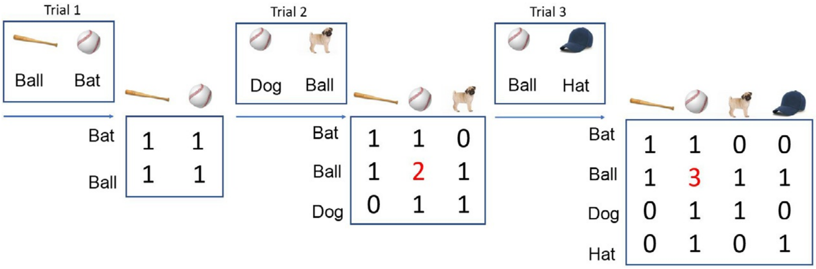


Fig. 4 Pursuit model.

2.2 Associative learning

An alternative mechanism for CSL is Associative Learning (AL), which has been proposed and supported by many empirical and modeling studies, suggesting that human learners are able to keep track of multiple possible word–object pairings simultaneously, and they use aggregated knowledge to inform later decisions (Frank, Goodman, & Tenenbaum, 2009; Smith & Yu, 2008; Vouloumanos, 2008; Yu, Ballard, & Aslin, 2005; Yu & Smith, 2007). In this model, word learning can be thought of as classic AL with multiple cues (i.e., objects) and outcomes (words). Words can be learned by accumulating information across situations, and this accumulating process is the key to all associative models of language acquisition. Different from the HT model discussed earlier, the AL model suggests that word–object association is not an all-or-none process. Instead, learning passes through a state of partial knowledge, and this partial stage is critical for building models and generating interesting predictions.

Fig. 5 shows how a simple AL model works using the same learning scenario described in Fig. 2. In the first learning scenario, learners would see two objects BALL and BAT simultaneously and hear two labels: "ball" and "bat." They do not know which word maps on to which object, creating a referentially ambiguous situation, but they know statistically there are four possible mappings. A cross-situational learner would keep track of all four mappings (BALL—"ball," BALL—"bat," BAT—"ball," BAT—"bat") and carry them over to the next learning trial. In the second trial, the learner would form four more mappings, but the object BALL and word "ball" cooccur again. With two counts of association value, this specific association now gets stronger than other associations. Similarly, in trial 3, the correct BALL—"ball" mapping is further strengthened with more cooccurrences. As learners encounter more and more learning situations,



Bat	1	1
Ball	1	1

Bat	1	1	0
Ball	1	2	1
Dog	0	1	1

Bat	1	1	0	0
Ball	1	3	1	1
Dog	0	1	1	0
Hat	0	1	0	1

Fig. 5 A toy example of associative learning.

eventually the correct mapping that the round object is called ball would get the strongest association because the label and its correct referent are likely to cooccur more consistently than do other pairs.

2.2.1 Empirical evidence on AL

Based on the CSL paradigm originally reported in Yu & Smith (2007), several studies have reported empirical evidence to support the AL approach (i.e., Smith, Smith, & Blythe, 2009; Smith et al., 2011). Moreover, recent studies using the HSP, described earlier, have also provided empirical support for the AL model. For example, Yurovsky, Smith, and Yu (2013) conducted a study using the same HSP used in Medina et al. (2011). In addition to recording parent–child natural interaction from a third-person perspective, like in the original HSP studies, they also recorded children's first-person view when naming occurred. They showed adult participants a series of 5-s naming events from either the first- or the third-person view and instructed the participants to guess the referent that the parent labeled. The result showed that guess accuracy varied considerably across vignettes. Based on the identification accuracy, about half of the naming instances were either highly ambiguous or highly unambiguous. Yurovsky et al. (2013) then tested whether learners could extract useful information from highly ambiguous naming instances. They showed participants four ambiguous vignettes in random orders and asked participants to make a guess after each video. Interestingly, there was a significant learning improvement across instances from the first-person view. Learners' accuracy increased from 12% at trial 1 to 26% at trial 4. However, this incremental learning pattern was absent from the third-person view. Trial 1 (10%) and trial 4 (15%) accuracies were not significantly different. In addition, participants presented with the first-person child view vignettes not only made a significant progress after a correct initial guess, but also after an incorrect initial guess, which contradicts the HT model suggesting that learning would only occur after initial successful guesses. Yurovsky et al.'s (2013) result is consistent with the AL view predicting that word learning does not only emerge from highly informative learning events, but also from aggregating information from less informative instances.

In a follow-up study using the same set of stimuli, Zhang, Yurovsky, and Yu (2015) conducted four experiments using the HSP to explicitly test whether adult learners can form multiple associations and carry over past knowledge to find potential word meanings or they carry only one conjecture forward and either confirm or disconfirm in the very next learning exposure (Propose-but-Verify). Zhang et al. (2015) examined in detail the mechanism by investigating how learning accuracy changed across learning instances to see how learning unfolds over time with accumulated experience. An analysis of the sequence of responses across learning instances revealed a better sense of what strategy learners tend to adopt. For example, if learners made an initial wrong guess that could not be confirmed in subsequent exposures, will they show baseline or improved performance in subsequent trials? Improved accuracy in learning outcomes would support the AL account and challenge the HT account, as the former suggests that learning is a continuous process for which all information, even ambiguous, could facilitate learning. In contrast, the HT account suggests that learners only store a few hypotheses and then reset the learning process if that hypothesis cannot be confirmed. According to this account, no improvement in accuracy would be seen after a hypothesis is rejected.

Zhang et al. (2015) found that when viewing two consecutive ambiguous trials, participants' current trial accuracy (39%) was still significantly above baseline (11%) even when they failed to find the correct target from the previous trial. More interestingly, as participants' total number of previous correct trials increased, their performance on the current trial also significantly improved, even in cases when their most immediate previous trial was wrong. This finding contradicts the Propose-but-Verify model, which suggests baseline accuracy after a wrong guess. These learning patterns demonstrate that learners did use their previous knowledge to guide current decision making and not just the immediate previous learning trials, but *all* previous experiences contributed to learning. Real-time behavioral data from this study revealed that learners can gradually accumulate knowledge from multiple word-learning situations with high uncertainty and that word learning is more likely to be a slow and continuous process rather than a one-shot, fast mapping one (Zhang et al., 2015).

2.2.2 Modeling work on AL

Researchers have built many computational models to understand how learning unfolds over time (e.g., Blythe, Smith, & Smith, 2010; De Beule, De Vylder, & Belpaeme, 2006; Fazly, Alishahi, & Stevenson, 2010; Fontanari & Cangelosi, 2011; Räsänen & Rasilo, 2015; Vogt, 2012). For example, using infant looking time data from a CSL task, Yu and Smith (2011) built a computational model with a goal to investigate the structure in looking patterns generated by the infants and whether certain looking patterns can predict learning performance at test. In the behavioral part of the study, Yu and Smith (2011) conducted a classic CSL task by providing infants with the opportunity to learn 6 novel words after experiencing 30 training trials. Each trial contains two novel words and objects. The two objects were presented simultaneously and were shown for 4 s. Following training were 12 testing trials. Infants saw two objects (one target, one distractor from the training set) and heard one word, repeated four times. Infants' looking behavior was tracked throughout training and testing and then subjected to an AL model.

To build an AL model using infant gaze fixation data, a 6 × 6 association matrix was used, as shown in the toy example in Fig. 6. With words listed on the *y* axis and objects listed on the *x* axis, the 2 × 2 matrix included all the possible associations that learners could possibly track at each trial. Each cell represented one specific word–object mapping. While the diagonal cells indicated correct mappings, nondiagonal cells represented mismatches due to ambiguity inherent in the task. Association strength was defined by the duration of fixations on that particular object when a label was uttered. For example, as shown in trial 1, if participants spent 70% of time looking at object BAT (represented by color black) and 30% of time looking at object BALL (represented by color white) while hearing the label "bat," the association strength for BAT—"bat" mapping would be 0.7 and the BALL—"bat" mapping would be 0.3. Similarly, if they looked at BAT 40% of time and BALL 60% of time while hearing the label "ball," the association strength for the BAT—"ball" mapping and BALL—"ball" mapping would be 0.4 and 0.6, respectively. The model tracks association probabilities, updating them trial by trial. By adding matrices from individual trials together, the model generated an accumulated matrix with association values created

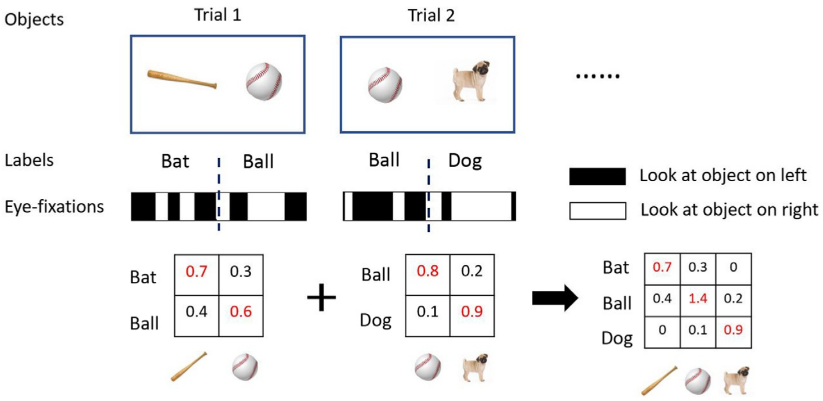


Fig. 6 A toy example of the associative learning model.

by trial-by-trial looking data from the entire training session. The model assumes that learners make final decisions during testing by selecting the strongest association from the learners' internal association matrix. Therefore, a successful learner's internal association matrix will have the highest association strength on diagonal cells.

Yu and Smith (2011) found that by aggregating statistical information across all learning trials, this model predicted infants' performance at test, indicating a strong correlation between predicted results from the model and the behavioral results from human learners. In addition, this model was sufficient to predict individual differences and differentiate strong and weak learners purely based on their looking patterns. This simple model was tested with data from the specific eye movement patterns made by infants. In other words, it is based on the specific associations infants formed and accumulated over trials. It is the case that learners, especially infants and young children, might not be able to store all of the word–object cooccurrences viewed due to their limited cognitive capacity. However, by being able to look at the right referent at the right moment in time seems to be a good predictor of a successful word learner. By investigating how learners selectively attend to information through the fine-grained analysis of their looking patterns to build word–object mappings, Yu and Smith (2011) showed that a simple AL model can provide a plausible model for cross-situational word learning.

In this section, we reviewed some past and current research investigating two fundamentally different word-learning models: HT and AL. By studying both adults and infants, researchers have found behavioral and computational evidence supporting each model. The disagreement from previous studies may be explained by the various methods and designs used in different studies, as they may rely on different assumptions regarding early word learning. To fully understand the learning mechanism underlying this debate on whether word learning follows a "fast mapping" procedure or a gradual statistical one, we will need to have a deeper understanding of different psychological components in the learning system.

3. Examining psychological components in a learning system

Instead of comparing a HT model with an AL model as a whole, one useful way to further compare the two is to decompose each model into psychological components. Then we can closely examine those components to see if there are differences and similarities between the two learning mechanisms. In the context of statistical word learning, we proposed partitioning a CSL model into three psychological components: the information selection processes on each trial; the learning machinery, which is usually viewed as the core of statistical learning; the decision processes used when participants are tested for their learning of word-referent correspondences (Yu & Smith, 2012).

3.1 Information selection

Cross-situational word-referent learning potentially solves the problem of learning the mapping of words to referents from individual word–scene correspondences that are inherently ambiguous. This is expected, in part, because real world scenes typically contain many objects and many likely referents, and also because multiple words may be used to talk about that scene at any instance. All this creates for the learner who does not yet know many word-referent pairs the problem of figuring out what word goes with what referent. The core idea of statistical word-referent learning is that the learner could solve the problem by combining word-referent cooccurrence data across trials.

But if there are many words and many potential objects, do learners notice and store them all? A key question for statistical learning is how much and what kind of information is selected and stored at each moment of learning. Even if one assumes that the units for learning are whole words (not their parts or phrases) and whole objects (not their parts, properties, or sets) and even if one limits the learning environment to that of laboratory cross-situational studies, there are still several words and several referents at each moment and thus potentially many different solutions to information selection. One could, as an ideal learner, register all the word-referent pairs at every learning moment; that is, all the possible pairings, consistent with the input at each moment, might be stored. To follow the toy example shown in Fig. 1, an ideal learner would store four word-referent pairs in the first trial (e.g., BALL—"ball," BAT—"bat," BAT—"ball," BALL—"bat"). Alternatively, and perhaps more consistent with what is known about human attention, one might attend to only a subset of words and referents, registering just partial information—some words, some referents—from all that is available at a single moment. Selection could be very narrow (e.g., just one word, one referent per learning moment) or it could be broader. Further, if learners do select just some of the information, what guides selection? It could be random and unrelated to past experience or it could be influenced by what they have seen before. In any case, the first step for a model of learning from a series of individually ambiguous learning trials is to specify the information input that the learning mechanism receives as we discuss next.

3.2 Learning machinery

In most discussions of statistical learning, the learning machinery is viewed as the central theoretical question. As discussed previously, there are two fundamentally different kinds of learning mechanisms and with fundamentally different implications about the nature of the learner. AL is characterized as the simple and uninformed registration of cooccurrences and/or the calculation of probabilities (Yu & Smith, 2007). HT, in contrast, is generally characterized as a form of learning in which coherent hypotheses are formed often in conceptually constrained ways and evaluated—not just by counting cooccurrences—but through some more rational evaluation of the evidence (Medina et al., 2011). Thus, there seems to be core differences between HT and AL. However, within these two classes, there are choices to be made about the learning mechanism itself. For example, an HT learner could keep track of and aggregate evidence for just some (but not all possible) word-referent hypotheses. If this is the case, the model needs to specify how those initial hypotheses are formed or selected as well as how many hypothesized pairs the learning system is capable of tracking. An HT model also needs to specify how strong the evidence needs to be for the learning system to accept or reject a hypothesized pair. An AL learner could simply count registered occurrences, or such a learner could apply more advanced probabilistic computations based on those counts (Kachergis et al., 2012), such as conditional probabilities (Aslin, Saffran, & Newport, 1998). In brief, both HT and AL models have choices about the kind of information aggregated, the kinds of computations applied to that information, and the form of the learning outcome.

3.3 Decisions at test

Finally, a learning system needs to specify how the accrued information is retrieved and used by learners to make decisions during testing. Commonly, experiments on word learning present the learner at test with a single word and some number of alternative referents (Yu & Smith, 2007; Yurovsky & Yu, 2008). Given the evidence accumulated during training—a list of hypotheses with weak or strong association strengths—participants must make a momentary decision, selecting the most likely referent given the queried word. Learners could apply a winner-take-all strategy such that the strongest hypothesis or association for the tested word governs choices in an all-or-none manner. Alternatively, responses could be graded and based on the probability that a hypothesis is correct or the relative strength of all the associations to the candidate word. Further, the decision could be based on the single word (or word-referent pairing) being queried or it could be based on a "best overall solution" for all the word–object pairings acquired during training. In brief, candidate-learning mechanisms also need to specify how acquired information is retrieved and how decisions are made at the time of test.

Considering the three psychological components (Information selection, Learning machinery, Decisions at test), recent simulation studies show that one type of model can simulate the other type of model (Yu & Smith, 2012). Depending on how these components interact, associative models can generate learning patterns that are created by hypothesis-testing models with fewer training trials. At the same time, the associative model can mimic various hypothesis-testing models by changing how these three components interact, producing the same learning patterns but through different learning processes. Specifically, Yu and Smith (2012) proposed a formal unified learning principle based on both learning mechanisms by identifying some common mechanisms that are shared between the two: (1) what information is stored on an individual learning trial, and (2) how stored information is evaluated and selected.

First, the AL model stores all possible cooccurrences in a big two-dimensional matrix, whereas the HT model only keeps a short list of hypotheses. The former takes in many cooccurrences probabilistically, whereas the latter only stores the most favored one and ignore all others. Yu and Smith (2012) suggested that the AL model can be converted into a HT model if we assign the mapping with the highest strength with 1 and all the others with 0. This makes HT a special binary case of AL. Even when learners store lexical knowledge in a probabilistic form, during testing, learners still need to select the most likely mapping and ignore all others. Different thresholds determine how many pairs learners keep but there is not a clear threshold to separate the two mechanisms.

Second, during the information retrieval stage, the two processes are also not fundamentally different. The lexical information stored in the association learning matrix can be activated in response to a specific input (e.g., hearing a word). To find the most likely referent, one can extract the strongest hypothesis set from the association matrix, or one can decompose the association matrix into several hypothesis sets and make final conjectures based on hypothesis test averaging. Yu and Smith (2012) concluded that there is a natural associative interpretation of the different strategies people may use and these strategies reside on a continuum. The main idea of this second point here is that the representational forms posited by associative and hypothesis-testing accounts of CSL may be understood as two cases of the same representation (Romberg & Yu, 2014). Although not a new idea (Mitchell, De Houwer, & Lovibond, 2009), it is not well recognized in the empirical domain of word-referent learning. It is an insight that may be particularly important to the understanding of developmental changes in word-referent learning. It has been suggested on the basis of empirical evidence that word learning proceeds from a more associationist beginning to more rapid and smart learning via HT (Hollich, Hirsh-Pasek, & Golinkoff, 2000). How that transition might happen and the nature of the processes that underlie it might be better understood if conceptualized in terms of a transition from denser and graded associationist matrices to sparser and binary ones. Such a conceptualization may also help us understand how adult performance sometimes looks like the simple accrual of cooccurrences and other times looks like computationally powerful HT (Rogers & McClelland, 2004). Admittedly, this is a theoretical conjecture and considerably more formal and analytic work that needs to be done. But it is an open possibility in need of systematic exploration.

In this section, we took a closer look at the two competing models using three directly comparable psychological components: (1) input/information selection (the amount of information, words, and objects learners choose to attend); (2) the learning and computational machinery; and (3) the retrieval and decision processes at test. Computational simulations using these components will likely yield testable but different predictions that each model would make regarding learning. These predictions could then be further tested behaviorally, which in turn will lead to more refined models that fit the data better. The goal of comparing components of learning systems is not to find which of the two word-learning models is better, but to help reconcile the tension between these two models and suggest how we can move from a heated debate to a mechanistic understanding of how learners systematically choose, process, and retrieve information from a learning system.

4. New directions

In behavioral studies, the choice of methods could greatly influence the type of data being collected and the findings drawn from those data (Tamis-LeMonda, Kuchirko, Luo, Escobar, & Bornstein, 2017). In recent years, developmental researchers have started to use state-of-the-art technology in inventing new methods to examine how infants break into words. These data capture a more comprehensive picture of infants' language experience and are likely to provide new insights on underlying word learning mechanisms. In this section, we will discuss three new research directions to investigate the mechanisms of CSL: (1) real-time behaviors; (2) real-world data; and (3) neuroimaging evidence.

4.1 Real-time behaviors

Looking behaviors have been primarily used as the outcome measures of infant word learning in many laboratory experiments: Infants "know" a name when they look to the correct referent upon hearing that name. Traditional experimental paradigms for word learning often use simple stimuli. Often times, learners see clear images displayed side by side on a monitor screen and hear labeling sentences that are loud and clear. However, everyday visual scenes are ambiguous with respect to the many likely referents of a heard name. One critical question is: given the visual scene information from the learners' perspective, how do learners visually select objects to attend to when object names are heard?

Selective attention is critical for all learning tasks. Learners need to be highly selective in sampling useful information from their environment to learn efficiently and effectively. One useful resource is to measure and examine visual attention in a CSL task by collecting real-time looking data over the course of learning as moment-by-moment gaze data can show how selective attention provides the foundation for building word–object associations. By capturing real-time looking dynamics, we will be able to better understand how learning unfolds over time as well as to compare how different learning strategies lead to different outcomes.

In the context of word learning, ideal learners with unlimited cognitive resources would have no problem finding the correct word–object mappings. However, it is impossible for human learners to store all cooccurrences due to limited attention and memory. If learners only select a subset of information to attend to, how much information and what kind of information do they select, process, and store? Furthermore, what factors guide information selection? Do learners randomly select objects to attend to or do past experiences of what they have previously seen, in the form of familiarity or novelty preferences, guide their attention?

To answers these questions, Yu, Zhong, and Fricker (2012) conducted an eye-tracking study with adults. They presented participants with a set of training trials, each containing four novel objects. Following exposure to the visual display, participants heard four novel word labels in sequence. Participants were then instructed to select the referent of each word after training. By analyzing participants' real-time gaze patterns, these researchers found that after hearing a word, participants dynamically allocated their attention across the four novel objects, and they most often spent a large proportion of time gazing on one object. Overall, at the individual word level, different learners exhibited similar patterns of visual attention in terms of the number of fixations and the durations of the longest looks when exposed to each object. At the within-trial level, one interesting pattern emerged was that participants tended to use mutually exclusive looking behaviors by looking at different objects during different word segments. This way, despite accuracy, each object received at least one look during each of the four-word segments. This pattern is particularly salient for learners who were more accurate at testing, suggesting that mutually exclusive looks may be a good strategy for information selection that contributes to successful learning. In addition, learners also used prior knowledge to structure their looks. After a word was considered learned, learners strategically arranged their looks based on the mutual-exclusivity constraint within a trial. Prior knowledge of correct mappings freed up more cognitive resources, allowing participants to apply novel labels to novel objects. This could reduce the degree of uncertainty within a trial to facilitate later learning.

Knowing how adults select information to learn word–object mappings, the next interesting question is: how do adults' visual attention patterns shed light on infant word learning? In the cross-situational word learning task, 12- to 14-month-old infants were able to track cooccurrences between labels and objects across trials and learn which label reliably cooccurs with each object (Smith & Yu, 2008). In a follow-up eye-tracking study using a similar CSL paradigm, Yu and Smith (2011) found that at the beginning of training, looking patterns were similar across all infants. As was the case with the adult participants' results, initial looks showed very little systematicity with many rapid short looks switching from one object to the other. This could be potentially helpful to word learning as it allows infants to sample data about referent objects more broadly. As training proceeded, infants who learned the word-referent mapping at test started to show different looking patterns compared to the nonlearners. Learners' gaze patterns became more systematic and gradually directed toward the correct referent (Yu & Smith, 2011). One advantage of using infant eye-tracking data is that this method goes beyond asking the question of whether or not infants can track distributional information by empirically testing how looking behaviors during training emerge across trials.

Relatedly, another approach to study infant visual attention is through computationally explicit models. One recent model of multiobject visual attention focused specifically on how attention shifts between objects in an ideal learning situation (Pelz, Piantadosi, & Kidd, 2015). The ideal learner model is trying to formalize what type of attentional behavior should be expected in a system that is aiming at gathering information efficiently from a complex environment. The parameters included in this process model were: (1) the learning curve that the learner follows; (2) memory decay rate, which sets limit on the learner's short-term memory capacity; (3) the cost of switching attention between objects; and (4) prior knowledge of the objects. At each step, the model would calculate the amount of information it expects to gain from each object and decide whether it should continue to attend to the same object or switch attention to another object based on the expected information gain. In a series of simulations, the model shows that once the system picks up an object, it will maintain that information held in memory by switching between all the previously attended objects. Once the maximum number of objects has been picked up, switching behavior becomes cyclical to maintain the amount of information remembered for each object. As decay rate increases, fewer objects can be attended to and the average amount of information learned for each object also decreases. Although the underlying dynamics of attention and learning are complicated, this process model provides several testable hypotheses that can be applied to human attentional systems.

Currently, we know very little about infants' visual attention, but the study of looking behaviors and changes in looking behaviors will provide detailed information with respect to understanding the mechanisms through which infants visually select objects in natural scenes and how that selection changes as a function of word and object-in-scene experiences.

4.2 Real-world data

Language input matters to learning outcomes in the real world and early language acquisition is based on data (Hart & Risley, 1995; Hoff, 2003; Weisleder & Fernald, 2013). Today, with the increasing computational ability to automatically process data, collecting data from naturalistic environments is getting more and more feasible and will likely contribute to developmental research in many ways. This approach will provide new directions for quantifying the quality of children's language input and revealing fresh perspectives on fundamental questions of language acquisition.

Data also matter to how learning models perform. Any statistical learning mechanism, either based on HT or AL, uses and relies on the input. Therefore, one critical aspect for studying infants' developing word-learning system is language input. Most studies on word learning are conducted in the laboratory using screen-based displays of stimuli. We do not really know what the child's natural visual experience is and how these visual inputs shape patterns of learning across the vast multitude of items/events in the real world and across multiple timescales. One challenge that young word learners face in early word learning is referential uncertainty (Quine, 1960). Many laboratory studies have been designed based on different assumptions on the degrees of uncertainty in real-world word learning. Some theorists assume infants' everyday learning environment is messy. Infants only learn at rare moments that are referentially transparent (Trueswell et al., 2013). Others assume that even though individual naming moment can be ambiguous, learners are able to use statistical learning across naming events to find word-referent pairings (Smith & Yu, 2008). Different assumptions have led to different kinds of experiments and theories. However, we should not be working from just assumptions alone. The ambiguity of the learning environment for young learners involves answerable empirical questions, just as: How likely are learners to be looking at the intended referent when it is named in everyday environments? Do the visual properties (e.g., size and color) of objects influence where they look when hearing a word label?

One way to study these questions is to combine a traditional screen-based eye-tracking paradigm with naturalistic stimuli. In a recent study by Zhang and Yu (2016), they explored the looking patterns of 12-month-old infants using naturalistic images with varying visual properties (i.e., big vs small object size, center vs off-center object location) in order to examine whether perceptual properties of objects in children's own view during naming moments would influence how young infants select candidate objects to build word–object mappings. They first measured the total number of objects attended to by infants and found that even given that there were, on average, more than 10 objects in view, and also given plenty of viewing time (7 s per image) to potentially attend to many objects, infants only selectively attended to 3–6 objects per trial. In addition, they examined how infants allocated their attention among the subset of objects they chose. Do infants attend to those objects equally frequently or do they only primarily attend to one or two objects? Zhang and Yu (2016) found that infants spent more than 50% of time looking at one selected object, suggesting that even though infants focused on only a few objects per trial, they predominantly only looked at one object at least half of the time. Their results indicate that visual properties of objects in infants' own view during naming moments directly impact how they select candidate objects to attend to. Information from the world is filtered through not only the dynamics of first-person views but also the learner's own developing attention system due to limited attentional resources.

A child's learning environment may not be as messy as we thought. Instead, there is a significant amount of information reduction through selective visual attention from the infant's own view. What is attended to by the infant is not everything in the world and is not even everything in their field of view. Infants' own visual worlds are going to incrementally change and influence how they select information and build word–object mappings. Using realistic input may provide new information on the ambiguity of the scenes that coincide with parent naming events in naturalistic environments, on how infants distribute gaze and visually select objects in those scenes, on how looking behaviors change as a function of the cooccurrences of heard words and visually selected objects, and on how this learning might culminate in the power of a heard word to direct gaze to the named object across many different scenes including both transparent and highly cluttered ones.

4.3 Neuroimaging evidence

With advancement in neuroimaging technology, one recent study done by Berens, Horst, and Bird (2018) collected functional MRI data during a cross-situational word learning task to investigate whether learners' neural representation of the learning process over time supports a gradual AL account or a rapid HT account. They found evidence demonstrating that the brain activities showed in the left hippocampus area support the Propose-but-Verify model, which further indicates that participants are more likely to form explicit hypotheses while performing the CSL task. The neuroimaging data are also consistent with participants' self-reported strategies, which are mostly related to HT.

This is the first piece of neuroimaging evidence collected using CSL task, but there are some limitations. As pointed out by Kachergis (2018), the task used in the study is fairly simple with many repetitions. Therefore, learners could be at different stages of learning, which may elicit different learning strategies. In addition, real-life situations are more likely to contain multiple word–object associations that make it challenging for learners to engage in explicit hypothesis-testing. This "messier" input could create a "partial knowledge" state, in which participants are not explicitly making one hypothesis, but storing multiple possible mappings with different association strengths with the correct mapping emerging over time. The neuroimaging results do not rule out the possibility that CSL could be built upon AL operating in the background (more implicit) and AL may just be a more general form of HT. More neuroimaging work on this topic is needed to help advance our understanding of how word learning unfolds at the neural level.

5. General discussions

This chapter reviewed some past and current experimental studies and computational models to elucidate aspects of the word learning process. Researchers largely agree that statistical learning is a powerful mechanism that learners can use to adapt to various features of the statistical structure of language. The main lesson from the experimental evidence reviewed here is that human learners do seem capable of using statistical information to learn word–object mappings from multiple data sources. Other than behavioral studies, to learn more about the abilities and biases of human learners, researchers need to keep investigating the step-by-step word learning processes by proposing different computational models and testing computer simulations.

The two classes of models (HT vs AL) might be considered to differ in the mere details of how cooccurrence data are used. That is, they differ in the operations that are performed on or the interactions within cooccurrence data. The problem for theorists is that those "details" (the potentially powerful operations) in how cooccurrence data are used may occur in a variety of ways and at several different points in the process of learning. Perhaps the best way to solve this quandary and to advance theory is to start with what we know to be the case and then ask how the machinery of AL might give rise to different patterns of outcomes that are reasonably described as HT. Modeling results help to make clear that we cannot judge which class of models best describes human performance without knowing more about the three separable steps of information selection, learning machinery and storage, and then the retrieval and use of that information to formulate a response in given tasks. Researchers need to better understand aspects of the model (from information selection to core machinery to decision making) and how it works, and we need to constrain those component processes by collecting empirical evidence from humans about those very same component processes.

Another critical point to keep in mind when investigating early language development is the importance of studying real-time behaviors in real-life contexts. Systematic lab experiments have long been the predominant approach to answer developmental questions about word learning. Although traditional methods have been very useful, the integration of the collection of naturalistic high-density data with real-time analyses of the cognitive processes involved in word learning will provide a valuable framework for generating research questions to gain a comprehensive understanding of children's lexical development.

Acknowledgments

This work was funded by National Institutes of Health Grant R01HD074601 and R01HD093792 to CY.

References

Akhtar, N., & Montague, L. (1999). Early lexical acquisition: The role of cross-situational learning. *First Language*, *19*, 347–358.

Aslin, R., Saffran, J., & Newport, E. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*, 321–324.

Aussems, S., & Vogt, P. (2015). Adults track multiple hypotheses simultaneously during word learning. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th annual conference of the cognitive science society* (pp. 2793–2798), Austin, TX: Cognitive Science Society.

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, *62*, 874–890.

Berens, S. C., Horst, J. S., & Bird, C. M. (2018). Cross-situational learning is supported by propose-but-verify hypothesis testing. *Current Biology*, *28*, 1132–1136.

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

Blythe, R. A., Smith, K., & Smith, A. D. (2010). Learning times for large lexicons through cross-situational learning. *Cognitive Science*, *34*, 620–642.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Proceedings of the Stanford Child Language Conference*, *15*, 17–29.

Chen, C. H., Zhang, Y., & Yu, C. (2018). Learning object names at different hierarchical levels using cross-situational statistics. *Cognitive Science*, *42*, 591–605.

De Beule, J., De Vylder, B., & Belpaeme, T. (2006). A cross-situational learning algorithm for damping homonymy in the guessing game. In *Artificial life X* (pp. 466–472). MIT Press.

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

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *58*. Serial No. 242.

Fitneva, S. A., & Christiansen, M. H. (2011). Looking in the wrong direction correlates with more accurate word learning. *Cognitive Science*, *35*, 367–380.

Fontanari, J. F., & Cangelosi, A. (2011). Cross-situational and supervised learning in the emergence of communication. *Interaction Studies*, *12*, 119–133.

Frank, M., Goodman, N., & Tenenbaum, J. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*, *20*, 578–585.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj, II (Ed.), *Vol. 2. Language development, language, thought and culture* (pp. 301–334). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Gleitman, L., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. (2005). Hard words. *Journal of Language Learning and Development*, *1*, 23–64.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H. Brookes Publishing.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*, 1368–1378.

Hollich, G., Hirsh-Pasek, K., & Golinkoff, R. M. (2000). I. What does it take to learn a word? *Monographs of the Society for Research in Child Development*, *65*, 1–16.

Kachergis, G. (2018). Word learning: Associations or hypothesis testing? *Current Biology*, *28*, R555–R557.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2012). An associative model of adaptive inference for learning word-referent mappings. *Psychonomic Bulletin & Review*, *19*, 317–324.

Koehne, J., & Crocker, M. W. (2015). The interplay of cross-situational word learning and sentence-level constraints. *Cognitive Science*, *39*, 849–889.

MacNamara, J. (1972). Cognitive basis of language learning in infants. *Psychological Review*, *79*, 1–13.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.

Markman, E. (1992). Constraints on word learning: Speculations about their nature, origins, and domain specificity. In M. R. Gunnar & M. P. Maratsos (Eds.), *Vol. 25. Minnesota symposium on child psychology* (pp. 59–101). Hillsdale, NJ: Erlbaum.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*, 121–157.

McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science*, *317*, 631.

Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 9014.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*, 183–198.

Monaghan, P., Mattock, K., Davies, R. A., & Smith, A. C. (2015). Gavagai is as gavagai does: Learning nouns and verbs from cross-situational statistics. *Cognitive Science*, *39*, 1099–1112.

Onnis, L., Edelman, S., & Waterfall, H. (2011). Local statistical learning under cross-situational uncertainty. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual meeting of the cognitive science society conference*.

Pelz, M., Piantadosi, S. T., & Kidd, C. (2015). The dynamics of idealized attention in complex learning environments. In *2015 Joint IEEE international conference on Development and learning and epigenetic robotics (ICDL-EpiRob)* (pp. 236–241), IEEE. August.

Quine, W. V. O. (1960). *Word and object (studies in communication)*. New York and London: Technology Press of MIT.

Räsänen, O., & Rasilo, H. (2015). A joint model of word segmentation and meaning acquisition through cross-situational learning. *Psychological Review*, *122*, 792–829.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. MIT press.

Romberg, A. R., & Yu, C. (2014). Interactions between statistical aggregation and hypothesis testing mechanisms during word learning. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th annual conference of the cognitive science society*. Austin, TX: Cognitive Science Society.

Scott, R. M., & Fisher, C. (2011). 2.5-Year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, *122*, 163–180.

Siskind, J. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*, 39–91.

Smith, K., Smith, A. D., & Blythe, R. A. (2009). Reconsidering human cross-situational learning capacities: A revision to Yu & Smith's (2007) experimental paradigm. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual conference of the cognitive science society* (pp. 2711–2716). Austin, TX: Cognitive Science Society.

Smith, K., Smith, A. D., & Blythe, R. A. (2011). Cross-situational learning: An experimental study of word-learning mechanisms. *Cognitive Science*, *35*, 480–498.

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

Stevens, J. S., Gleitman, L. R., Trueswell, J. C., & Yang, C. (2017). The pursuit of word meanings. *Cognitive Science*, *41*, 638–676.

Suanda, S. H., Mugwanya, N., & Namy, L. L. (2014). Cross-situational statistical word learning in young children. *Journal of Experimental Child Psychology*, *126*, 395–411.

Tamis-LeMonda, C. S., Kuchirko, Y., Luo, R., Escobar, K., & Bornstein, M. H. (2017). Power in methods: Language to infants in structured and naturalistic contexts. *Developmental Science*, *20*, e12456.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *57*, 1454–1463.

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*, 126–156.

Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, *127*, 375–382.

Vogt, P. (2012). Exploring the robustness of cross-situational learning under zipfian distributions. *Cognitive Science*, *36*, 726–739.

Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition*, *107*, 729–742.

Vouloumanos, A., & Werker, J. (2009). Infants' learning of novel words in a stochastic environment. *Developmental Psychology*, *45*, 1611–1617.

Wang, F. H., & Mintz, T. H. (2018). The role of reference in cross-situational word learning. *Cognition*, *170*, 64–75.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*, 2143–2152.

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

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

Yu, C., & Smith, L. B. (2011). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental Science*, *14*, 165–180.

Yu, C., & Smith, L. B. (2012). Modeling cross-situational word learning: Prior questions. *Psychological Review*, *119*, 21–39.

Yu, C., Zhong, Y., & Fricker, D. (2012). Selective attention in cross-situational statistical learning: Evidence from eye tracking. *Frontiers in Psychology*, *3*, 1–16.

Yurovsky, D., Smith, L. B., & Yu, C. (2013). Statistical word learning at scale: The baby's view is better. *Developmental Science*, *16*, 959–966.

Yurovsky, D., & Yu, C. (2008). Mutual exclusivity in cross-situational statistical learning. In B. Love, K. McRae, & S. VM (Eds.), *Proceedings of the 30th annual conference of the cognitive science society* (pp. 715–720). Austin, TX: Cognitive Science Society.

Zhang, Y., & Yu, C. (2016). Examining referential uncertainty in naturalistic contexts from the child's view: Evidence from an eye-tracking study with infants. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th annual conference of the cognitive science society* (pp. 2027–2032). Austin, TX: Cognitive Science Society.

Zhang, Y., Yurovsky, D., & Yu, C. (2015). Statistical word learning is a continuous process: Evidence from the human simulation paradigm. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th annual conference of the cognitive science society* (pp. 2793–2798). Austin, TX: Cognitive Science Society.