Published in final edited form as:
Trends Cogn Sci. 2014 May ; 18(5): 251–258. doi:10.1016/j.tics.2014.02.007.

The unrealized promise of infant statistical word-referent learning

Linda B. Smith, **Sumarga H. Suanda**, and **Chen Yu**
Psychological and Brain Sciences, Program in Cognitive Science, Indiana University, Bloomington, IN 47405

Abstract

Recent theory and experiments offer a new solution as to how infant learners may break into word learning, by using *cross-situational statistics* to find the underlying word-referent mappings. Computational models demonstrate the in-principle plausibility of this statistical learning solution and experimental evidence shows that infants can aggregate and make statistically appropriate decisions from word-referent co-occurrence data. We review these contributions and then identify the gaps in current knowledge that prevent a confident conclusion about whether cross-situational learning is the mechanism through which infants break into word learning. We propose an agenda to address that gap that focuses on detailing the statistics in the learning environment and the cognitive processes that make use of those statistics.

The world offers data to novice word learners in the form of word-object co-occurrences. These data may be quite noisy with many spurious co-occurrences [Figure 1]. Thus a core problem for theories of early word learning is determining how infants manage to find the right word-referent pairs in the noise. The evidence indicates that by their first birthday, if not before, infants have already found a considerable number of these correspondences [1-3]. Older word learners, two-year-olds, employ knowledge about social cues, language, and categories to map words to referents; however, this knowledge develops over the course of word learning and may be partly a product of word learning itself [4-8]. Thus, the field lacks an understanding of how early word learning gets its start. Recent theory and experiments offer a new solution: novice learners may break into word learning through the noisy co-occurrence data, using *cross-situational statistics* to find the underlying word-referent mappings. We begin with a review of the models that show the in-principle plausibility of this solution and then turn to the experimental evidence showing that infants aggregate and make statistically appropriate decisions from co-occurrence data. We then turn to the critical question: Could this solution work for infants in the real world? The answer depends on a better understanding than we currently have of the relevant statistics in the learning environment and the cognitive processes that make use of those statistics.

Corresponding author: Smith, L.B. (smith4@indiana.edu).

Finding Structure in Co-occurrences

The classic debate in the study of early word learning pits hypothesis testing against associative learning [6, 9]. Recent computational advances have blurred the distinction; models from both frameworks can operate over the same co-occurrence data to yield the same learning patterns [see Box 1]. Accordingly, we ignore this classic divide to focus on four new contributions.

1. There is discoverable structure in word-scene co-occurrence data

Several researchers have applied statistical learning algorithms to word-scene co-occurrence data taken from audio and video recordings of infants in common everyday activities (i.e., an infant interacting with a parent). The algorithms, expressed as either Bayesian inference models [10] or machine translation models [11-13], readily succeed in discovering the underlying word-referent pairings from real-world co-occurrences. Although these powerful models may not be psychologically realistic [9], they show that there *is* significant structure in the co-occurrence data such that the co-occurrences –along with statistical learning mechanisms– might be enough for infants to discover the mappings of words to objects.

2. Statistical learning is about learning a *system of* co-occurrences

Statistical learning models succeed because they operate on the *set* of co-occurrence data with the goal of simultaneously learning multiple words and referents. Within these models, the strength of an individual word-referent association (or the probability of a hypothesis) is not strictly a property of that word-referent pair alone. Instead, it interacts with and is embedded in the other regularities in the co-occurrence matrix, regularities that enable the learning machinery to exploit correlations [15, 18, 22], coherent-covariation [18, 23] and the structure in the whole matrix [17,19, 24] to discover the underlying correspondences.

3. Word-referent pairs compete

One mechanism through which cross-item dependencies can influence learning is competition. Most current statistical models of word-referent mappings are proposed to block or inhibit other mappings that contain overlapping components [9, 25]. Such competition may also be responsible for the disambiguation (also known as “fast-mapping”) phenomenon in young children [Figure 2]. However, within the models, competition may be implicit and operate on partial (not behaviorally evident) knowledge [25]. For infant learners, this competition could mean that strong evidence for some mappings effectively “cleans up” the data helping the learning of weaker and noisier (non-competing) contingencies.

4. Statistical learning models learn to learn

Infants become better word learners as they learn more words. This is evident in the *vocabulary spurt* [14], an increase in the rate of receptive and productive vocabulary that typically emerges as infants approach their second birthday, as well as in a number of other emergent phenomena that are predicted by, and predictive of, infants’ vocabularies [Figure 2]. A number of models, including both associationist and probabilistic inference ones, have shown how these properties of early word learning may be driven by the statistics of word-referent co-occurrences [14 – 22]. Thus, statistical learning –and the structure in word-referent co-occurrence data–might not just start infant word-referent learning but might also build the knowledge-based reduction of uncertainty that is evident in older word learners.

In sum, the extant models serve as compelling demonstration proofs: They show that there is structure in the co-occurrence data. They shift the empirical question from one about how infants learn individual word-referent pairs to how infants operate on the statistical regularities within a system of words and referents to simultaneously learn multiple words. They highlight the potential importance of mutual exclusivity and competition among word-referent pairings as critical to the process and they yield developmental patterns of learning that are consistent with patterns observed in infant learners. With a few exceptions they do not attempt to model the real-time psychological processes –attention, memory--essential to actually learning words (9, 15).

Infants Aggregate Co-occurrence Data

The cross-situational word-learning task was invented to answer the question of whether infants can do what the models propose: learn multiple word-referent pairs from noisy co-occurrence data. On each trial in the task [Figure 1B], infants hear two words and see two objects with no information about which word goes with which object. However, across trials each individual word always co-occurs with just one referent; thus there is across trial certainty despite within trial uncertainty, *if* learners aggregate the co-occurrence data across trials. Smith and Yu [26; replicated 27] presented 12- and 14-month-old infants with a randomly ordered stream of 30 such trials with six novel words and six novel objects. At the end of this experience, infants' word learning was measured using a two-alternative preferential looking procedure: two visual objects were presented in the context of one spoken word and looking time to the statistically correct referent of that word was measured. The results showed that the infants looked more to the correct referent than the foil upon hearing the associated name. Moreover, analyses of individual word pairs and individual infants suggested that most infants learned four or more word pairs. To do this, infants must have attended to, stored, and in some way statistically evaluated the system of word-referent co-occurrences. In a subsequent study, Vouloumanous and Werker [28] showed that 18-month-old infants tracked both high and low frequency contingencies within a data set of word-to-object co-occurrences, a result consistent with the notion that infants track a system of regularities, not just the most dominant ones.

Two additional studies revealed that infant statistical learning is constrained by developing attention and memory processes. Smith and Yu [29] disrupted word-referent learning in 12- to 14-month-old infants by setting a novelty trap such that within each learning trial one potential referent was more salient than the other. They did this by ordering the training trials [Figure 1C] so that one object (and its location) was repeated within blocks of 5 trials while the other location showed a new object on each trial within that block of trials. Across the 30 trials, each object served equally often as the constant object and as one of the non-repeating objects, so that the final co-occurrence statistics were the same as in Yu and Smith's initial [26] study. This manipulation caused about half the infants to visually attend to only the non-repeating object within a trial; these infants failed to learn the underlying word-referent correspondences. In contrast, the infants who more successfully negotiated the novelty trap so as to sample both potential referents within a trial learned the underlying structure of word-referent mappings. Past research on visual attention indicates that infant attention is often too stimulus driven and too sticky such that infants may not fully sample the available information in an array [30,31]. Thus, the immature attentional system could pose a serious limit on infant statistical learning as the data available to any statistical learning machinery are not the data in the real world, but only the subset of that data that makes contact with the infant's learning system. Past research has also shown considerable individual differences among infants in the development of visual processing and specifically that more flexible visual vocabulary development [32]. Thus, the development of attentional processes that support a broad sampling of the co-occurrence data could be a rate-limiting factor in early vocabulary development [see also, 33].

In a related study, Vlach and Johnson [34] re-ordered the sequence of learning trials to test the role of memory in statistical word-referent learning. They arranged the series of trials [Figure 1D] so that some word-referent pairs were repeated on successive trials and others were only repeated after many intervening trials. They found that 20-month-old infants learned both classes of pairs, indicating that they could combine information over relatively many interleaved items. In contrast, 16-month-old infants only learned the successively repeated pairs, raising the question of whether the memory systems of young infants are sufficient to aggregate information across long delays between encounters of a word and referent. Critically, the statistics in all these experiments are likely to be different from those that characterize infant learning environments. Word-referent pairings in real world experiences are likely to have a bursty structure [35], with a word-referent pair repeating multiple times in a conversational context followed by periods where it is rarely encountered, and then by another burst of repetitions in another coherent conversation. How might this dynamic structure – and the cumulative repetitions over days and weeks – interact with the aggregation of information over time and with the statistical evaluation of that information? Evidence on the dynamic properties of co-occurrences in the world and evidence on how these properties engage the attentional and memory systems of infants is critically needed to determine the scalability of cross-situational word learning [Box 2]. The models and the infant experiments suggest cross-situational word learning as a potential mechanism for early word learning. However, the models provide computational rather than process accounts of learning. Furthermore, the models do not specifically model infant cross-situational learning data. The infant cross-situational learning task presents streams of word-object co-occurrences that are considerably simpler than those in the world and even so success is limited in some presentation arrangements for learners by their developing attentional and memory processes. Thus, the key question is still open: Could cross-situational word-referent learning work in the real world given the statistical regularities in infants' learning environments and given the seemingly limited cognitive skills of infants at the young ages at which initial word-referent learning occurs?

One on-going discussion in the literature centers on the question of just how noisy the data are. As a counterweight to the perhaps too powerful statistical learning algorithms, Trueswell, Gleitman and colleagues [36, 37] attempted to measure the uncertainty of word-referent co-occurrences in infant directed speech via what they call the "Human Simulation

Paradigm" (HSP). Instead of putting word-referent co-occurrences into a model, they presented adults with brief video clips of a parent talking to an infant. The audio was removed and a sound was inserted where the target word had been. The adults' task on each trial was to explicitly offer a hypothesis as to the intended referent by the parent at the moment of the cued sound. Adults were very poor at this and showed no ability to aggregate information about word-referent correspondences across trials. The researchers concluded that the co-occurrences in the real world are much too variants of the HSP method have come to different conclusions. One study [38] found that about 50% of the naming episodes by mothers to toddlers were not ambiguous to the adults who could readily guess the target referent. Another study [39] showed that the degree of word-referent ambiguity as measured by the HSP varied considerably across parents and that this ambiguity was predictive of later word learning, such that toddlers who experienced less ambiguity went on to develop larger vocabularies. Finally, in an ideal observer analysis of word-referent co-occurrences in parents' interactions with their 6- to 18-month-old infants, Frank and colleagues [35] found that parents mostly create unambiguous naming moments. If much of the early data are relatively clean, infants could simply ignore moments that are too uncertain [40] and use relatively simple statistical mechanisms to determine the word-referent pairs from this cleaner sample of co-occurrence data.

To date the methods for measuring the structure in the learning environment have taken word-referent co-occurrences from the world (videos of parent naming events) and submitted these co-occurrences to powerful algorithms or adult coders, analyzing the problem at the computational level. There are two limits to this approach. First, the data relevant to a statistical learner are not the data in the world but the data in the world filtered through the learners' sensory, attentional and memory systems [Box 3]. Second, these filters – sensory, attentional, and memory systems - are themselves statistical learners [e.g., 41-43] that may not simply let information through to the learner but may also weight that information in ways directly linked to statistical computation and to the co-variation in the data. If these ideas are right, we will need to change the very way we statistical regularities in the learning environment and we need to study how statistics are filtered and aggregated in multiple mechanisms from attention, to memory, to decision processes.

We may also need to ask what is the signal and what is the noise in the co-occurrence data. Consider again the stream of learning experiences illustrated in Figure 1A: In terms of the depicted co-occurrence matrix, these are very noisy data with many "spurious" correspondences. However, in another sense these are not spurious correspondences but examples of the coherent covariation that characterizes learning environments: for infants, spoons (and their name) may be typically experienced in contexts of oatmeal and sippy cups (and their names). What is already known about visual statistical learning [42, 42], about cued attention [44, 45], about the priming of memories [46], about the statistical structure of language [8] – is that this structure may actively help learners find the right correspondences. For example, the word "bowl" (or the sight of a bowl) might predict the likely presence of spoons and enable the learner to find that referent in the visual clutter; the word "bowl" (or sight of a bowl) may prime and activate memories of spoons and their names, thereby supporting the aggregation of information over time and contexts.

Theories of infant word-referent learning treat the co-occurrences between to-be-learned words and referents as the signal and all else as noise. But from another perspective, the “noise” contains information, regularities that interact with the sensory, attentional and memory processes on which cross-situational learning depends. It may be through the interactions of multiple statistically sensitive processes that novice learners simultaneously solve multiple and mutually constraining tasks of mapping words to referents while building semantic networks [47-50]. This proposal sets a possible agenda for future research.

Acknowledgments

This work was supported by NIH R01HD056029 to CY and LBS and NSF SMA-1203420 to SHS.

References

1. Bergelson E, Swingley D. At 6–9 months, human infants know the meanings of many common nouns. Proc Nat Acad Sci. 2012; 109:3253–3258. [PubMed: 22331874]
2. Tincoff R, Jusczyk PW. Six-month-olds comprehend words that refer to parts of the body. Infancy. 2012; 17:432–444.
3. Mani N, Plunkett K. Twelve-month-olds know their cups from their keps and tups. Infancy. 2010; 15:445–470.
4. Tomasello, M.; Tomasello, M. Constructing a language: A usage-based theory of language acquisition. Harvard University Press; Cambridge: 2009.
5. Smith LB, et al. Knowledge as process: contextually cued attention and early word learning. Cognitive Science. 2010; 34:1287–1314. [PubMed: 21116438]
6. Waxman SR, Gelman SA. Early word-learning entails reference, not merely associations. Trends in Cognitive Science. 2009; 13:258–263.
7. Bion RA, et al. Fast mapping, slow learning: Disambiguation of novel word-object mappings in relation to vocabulary learning at 18, 24, and 30 months. Cognition. 2012; 126:5339–53.
8. Romberg AR, Saffran JR. Statistical learning and language acquisition. Wiley Interdisciplinary Reviews: Cognitive Science. 2010; 1:906–914. [PubMed: 21666883]
9. Yu C, Smith LB. Modeling cross-situational word–referent learning: Prior questions. Psychol Rev. 2012; 119:21–39. [PubMed: 22229490]
10. Frank MC, et al. Using speakers’ referential intentions to model early cross-situational word learning. Psychol Sci. 2009; 20:579–585.
11. Yu C, et al. The Role of embodied intention in early lexical acquisition. Cognitive Science. 2005; 29:961–1005. [PubMed: 21702799]
12. Yu C, Ballard DH. A unified model of early word learning: Integrating statistical and social cues. Neurocomputing. 2007:702149–2165.
13. Yu C. A statistical associative account of vocabulary growth in early word learning. Language Learning and Acquisition. 2008; 4(1):32–62.
14. McMurray B. Defusing the childhood vocabulary explosion. Science. 2007; 317:631–631. [PubMed: 17673655]
15. McMurray B, et al. Word learning emerges from the interaction of online referent selection and slow associative learning. Psychol Review. 2012; 119:831–877.
16. Fazly A, et al. A probabilistic computational model of cross-situational word learning. Cognitive Science. 2010; 34(6):1017–1063. [PubMed: 21564243]
17. Kachergis G, et al. An associative model of adaptive inference for learning word-referent mappings. Psychon Bulletin & Rev. 2012; 19:317–324.
18. Colunga E, Smith LB. From the lexicon to expectations about kinds: a role for associative learning. Psychol Rev. 2005; 112(2):347. [PubMed: 15783290]
19. Xu F, Tenenbaum JB. Word learning as Bayesian inference. Psychol Rev. 2007; 114:245. [PubMed: 17500627]

20. Regier T. The emergence of words: Attentional learning in form and meaning. Cognitive Science. 2005; 29:819–865. [PubMed: 21702796]
21. Hidaka S, Smith LB. Packing: a geometric analysis of feature selection and category formation. Cognitive Systems Research. 2011; 12:1–18. [PubMed: 21442046]
22. Li P, et al. Dynamic self-rganization and early lexical development in children. Cognitive Science. 2007; 31(4):581–612. [PubMed: 21635309]
23. McClelland JL, et al. Letting structure emerge: connectionist and dynamical systems approaches to cognition. Trends in Cognitive Sciencs. 2010; 14:348–356.
24. Tenenbaum JB, et al. How to grow a mind: Statistics, structure, and abstraction. Science. 2011; 331:1279–1285. [PubMed: 21393536]
25. Yurovsky D, et al. Competitive processes in cross-situational word learning. Cognitive Science. 2013; 37:891–921. [PubMed: 23607610]
26. Smith L, Yu C. Infants rapidly learn word-referent mappings via cross-situational statistics. Cognition. 2008; 106:1558–1568. [PubMed: 17692305]
27. Yu C, Smith LB. What you learn is what you see: Using eye movements to study infant cross-situational word learning. Devel Sci. 2011; 14:165–180. [PubMed: 22213894]
28. Vouloumanos A, Werker JF. Infants’ learning of novel words in a stochastic environment. Devel Psychol. 2009; 45:1611–1617. [PubMed: 19899918]
29. Smith LB, Yu C. Visual attention is not enough: individual differences in statistical word-referent learning in infants. Language Learning and Development. 2013; 9:25–49.
30. Colombo J. On the neural mechanisms underlying developmental and individual differences in visual fixation in infancy: Two hypotheses. Devel Rev. 1995; 15(2):97–135.
31. Oakes, LM. IInfant perception and cognition: Recent advances, emerging theories, and future directions. Oxford University Press; 2011.
32. Ellis EM, et al. Visual prediction in infancy: What is the association with later vocabulary? Language Learning and Development. 2013; 10:36–50.
33. Fitneva SA, Christiansen MH. Looking in the wrong direction correlates with more accurate word learning. Cognitive Science. 2011; 35:367–380. [PubMed: 21429004]
34. Vlach HA, Johnson SP. Memory constraints on infants’ cross-situational statistical learning. Cognition. 2013; 127:375–382. [PubMed: 23545387]
35. Frank MC, et al. Social and discourse contributions to the determination of reference in cross-situational word learning. Language Learning and Development. 2013; 9:1–24.
36. Medina TN, et al. How words can and cannot be learned by observation. Proc Nat Acad Sci. 2011; 108:9014–9019. [PubMed: 21576483]
37. Trueswell JC, et al. Propose but verify: Fast mapping meets cross-situational word learning. Cogn Psychol. 2013; 66:126–156. [PubMed: 23142693]
38. Yurovsky D, et al. Statistical word learning at scale: the baby’s view is better. Devel Sci. 2013; 16:959–966. [PubMed: 24118720]
39. Cartmill EA, et al. Quality of early parent input predicts child vocabulary 3 years later. Proc Nat Acad Sci. 2013; 110:11278–11283. [PubMed: 23798423]
40. Kidd C, et al. The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. PLoS One. 2012; 7(5):e36399. [PubMed: 22649492]
41. Turk-Browne NB, Scholl BJ. Flexible visual statistical learning: transfer across space and time. J Exp Psychol: Hum Per Perf. 2009; 35:195.
42. Alvarez GA. Representing multiple objects as an ensemble enhances visual cognition. Trends in Cognitive Sciences. 2011; 15:122–131. [PubMed: 21292539]
43. Brady TF, Alvarez GA. Hierarchical Encoding in Visual Working Memory Ensemble Statistics Bias Memory for Individual Items. Psychol Sci. 2011; 22:384–392. [PubMed: 21296808]
44. Patai EZ, et al. Long-term memories bias sensitivity and target selection in complex scenes. J Cog Neur. 2012; 24:2281–2291.
45. Summerfield C, Egner T. Expectation (and attention) in visual cognition. Trends in Cognitive Sciences. 2009; 13:403–409. [PubMed: 19716752]

46. Squire LR. Memory systems of the brain: a brief history and current perspective. Neurobio Learning Mem. 2004; 82:171–177.
47. Thiessen ED. Effects of visual information on adults' and infants' auditory statistical learning. Cognitive Science. 2010; 34:1093–1106. [PubMed: 21564244]
48. Yeung HH, et al. Referential labeling can facilitate phonetic learning in infancy. Child Dev. 2013 Early View Online. DOI: 10.1111/cdev.12185.
49. Hay JF, et al. Linking sounds to meanings: infant statistical learning in a natural language. Cogn Psychol. 2011; 63:93–106. [PubMed: 21762650]
50. Hills TT, et al. The associative structure of language: Contextual diversity in early word learning. J Mem Lang. 2010; 63(3):259–273. [PubMed: 20835374]
51. Tulving, EE.; Craik, FI. The Oxford handbook of memory. Oxford University Press; 2000.
52. Perry L, et al. Highchair philosophers: the impact of seating context-dependent exploration on children's naming biases. Devel Sci. 2013 Early Online View, DOI: 10.1111/desc.12147.
53. Vlach HA, Sandhofer CM. Distributing learning over time: The spacing effect in children's acquisition and generalization of science concepts. Child Devel. 2012; 83:1137–1144. [PubMed: 22616822]
54. Vlach HA, Sandhofer CM. Developmental differences in children's context-dependent word learning. J Exp Child Psychol. 2011; 108:394–401. [PubMed: 21074169]
55. Werchan DM, Gómez RL. Wakefulness (not sleep) promotes generalization of word Learning in 2.5-year-old children. Child Devel. 2013 Early Online View, DOI: 10.1111/cdev.12149.
56. Yoshida H, Smith LB. What's in view for toddlers? Using a head camera to study visual experience. Infancy. 2008; 13(3):229–248. [PubMed: 20585411]
57. Smith LB, et al. Not your mother's view: The dynamics of toddler visual experience. Devel Sci. 2011; 14:9–17. [PubMed: 21159083]
58. Aslin RN. How infants view natural scenes gathered from a head-mounted camera. Optom Vis Sci. 2009; 86:561–565. [PubMed: 19417702]
59. Yu C, Smith LB. Embodied attention and word learning by toddlers. Cognition. 2012; 125:244–262. [PubMed: 22878116]
60. Pereira A, et al. A bottom-up view of toddler word learning. Psychon Bull Rev. 2013:1–8. Early View Online, DOI 10.3758/s12423-013-0466-4. [PubMed: 23090749]
61. Franchak JM, et al. Head-mounted eye tracking: A new method to describe infant looking. Child Devel. 2011; 82:1738–1750. [PubMed: 22023310]
62. Yu C, Smith LB. Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. PloS One. 2013; 8(11):e79659. [PubMed: 24236151]
63. Kretch K, et al. What infants see depends on locomotor posture. Journal of Vision. 2012; 12:182–182.
64. James KH, et al. Young children's self-generated object views and object recognition. J of Cogn Devel. 2014 in press.

BOX 1

Hypothesis Testing and Associative Learning

A the computational level, models of statistical word-referent learning may be characterized and formulated in terms of three components as in Panel A below: assumptions about the co-occurrence data on which the learning system operates, the statistical computations that are performed, and the learning outcomes that are achieved. The proposed statistical computations are generally seen as the main claim being made by the model. Within associative theories, these computations emerge from the strengthening and weakening of associations as a function of co-occurrence strength and competition among associations. Within hypothesis testing theories, conceptually coherent hypotheses are confirmed or disconfirmed through a variety of statistical procedures. These two frameworks thus offer fundamentally different characterizations of what it means to be a statistical learner.

But to actually simulate performance in a task, models of both classes must (implicitly or explicitly) specify a number of separable processing components as illustrated in Panel B. First, they must specify the *information selected* within a learning moment. The information selected within a learning trial could be narrow (just one word and one referent per learning moment), broad (many co-occurring word-referent pairs), or even change with learning (beginning broad and then narrower as more is known). Many or few hypothesized or association pairs may be stored in memory and thus tracked. As part of the statistical computation, models must specify how the tracked information is aggregated and represented, including for example, whether that *stored information* (knowledge) is represented in an all-or-none or probabilistic fashion. Finally, to simulate the learning outcome, the model must also specify the *decision processes* at test, including how information is retrieved. For example, the participant may make all- or-none decisions from graded statistics or make graded responses from those same data, and these decision processes, from the very same represented knowledge, could vary with testing context.

In a series of simulations, Yu & Smith [9] showed that these component decisions interact in complex ways within both classes of associative-learning and hypothesis-testing models. Indeed, very simple associative models could mimic sophisticated hypothesis testing models, producing the same learning patterns although different internal components. In brief, the two classes of models cannot be discriminated when formulated at the computational level in Panel A but only by direct assessments of the proposed learning mechanism in the context of explicitly specified supporting processing components as shown in Panel B. And all of should be informed by empirical evidence on infant attention and memory systems, as well decision processes.

BOX 2

Statistics in Time

Recent studies of infant and toddler retention of an encountered word and referent, suggest that learning is incremental and slow. For example, toddlers who map the novel word to the novel object in the fast-mapping paradigm in Figure 2, do not retain that mapping when tested after delays as short as 5 minutes, though they show savings in later learning [15]. Learning that is incremental and aggregates over time requires that the learning system recognize when an item is a repetition of previously experienced instance. For simplicity, consider the case, illustrated below, in which there is no within-situation ambiguity: each moment in time presents the learner with one object and one word, as illustrated below. How does the learning system know to aggregate over instance 1 and 4, that this is a “repetition?” Research on human memory [51] indicates three stimulus dimensions relevant to the likelihood with which previous memories are activated and combined with current input: similarity, time and context. For example, if the same person says “cup” across two instances (rather than if a male speaker names one and a female male, female talkers) or if the two cups are identical rather than perceptibly different, then the second naming event is more likely to activate and thus strengthen the memory of the first naming event. Likewise, a second instance that follows the first close in time is more likely than one that follows after a delay to activate and be combined with the memory of the first. Finally, human memory systems are highly dependent on contextual cues for activating memories. Thus, naming of a spoon in the context of a cereal bowl on a highchair tray is more likely to reactivate prior memories of spoon naming events that also occurred in that context than would be the naming of a spoon that is in the flatbed of a toy truck [52]. Similarity, time and context have all be demonstrated to play a role in toddler memory and retention of object names [53, 54] with delays (and partial forgetting) playing a positive role in building more abstract memories and generalizable knowledge [55] but with the exception of the Vlach & Johnson [34] have not been studied in the context of cross-situational word learning in the laboratory. The structure of the statistics of natural learning experiences along these dimensions also have not been studied.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

BOX 3

Three Lessons from the Infant's View of Word Learning

Developmental scientists have begun to take an "infant's perspective" on word learning by placing mini lightweight head cameras [56-60] or head-mounted eye trackers [61-63] on infants as they engage in various tasks. These new methods have documented unique properties of infant views. The growing body of work employing this technique has led to significant advancements in our understanding early natural vision [58], motor [61, 63], social [62], object processing [64], and language [10, 59, 60] development.

For scholars of early word learning, the existing findings from infants' views have offered at least three important lessons. First, an analysis of infants' views raises the possible need to reconsider fundamental assumptions about the problem of early word learning. That is, multiple studies have demonstrated that the infants' views of events are significantly *less cluttered* than adults' views of the same events [1-2,4]. This insight is significant considering the assumption that there is a high degree of referential uncertainty caused by cluttered everyday environments [11]. Second, an analysis of infants' views during naming events highlights the potential role of various visual properties in early word learning. In a series of studies in which parents played with and talked about novel objects with their toddlers, Yu, Smith and their colleagues observed that the visual properties of the target object during naming events (e.g., its image size relative to other objects, its centeredness in the visual field) predicted toddlers' novel object name learning [4-5], suggesting a possible larger role for bottom-up cues to word learning than previously suggested. This perceptual information may be directly linked to internal statistical computations. Finally, the information available in infants' views during naming events may actually be more conducive to and induce particular learning mechanisms. In a recent study, Yurovsky and colleagues [38] found that when naming events were viewed from an infant perspective (as opposed to a third-person perspective), adult word learners were more likely to engage in an aggregative, statistical word learning process. Together, these findings illustrate the need to take the infant's views into account – statistical information that makes contact with the infant's learning system -- when developing mechanistic theories of early word learning.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Outstanding Questions

What are the linguistic regularities in early naming events –isolated words, frequent frames, co-occurring object names?

What are the visual regularities in early naming events – visually isolated objects, saliency properties, co-occurring objects and contexts?

What are the regularities across words and objects and visual contexts?

What are the dynamic properties of repeated naming events – within the seconds and minutes of working memory processes, within the hours, days, and weeks of infant learning experiences?

Can these linguistic and visual regularities in the real world be represented in the framework of cross-situational learning?

NIH-PA Author Manuscript

NIH-PA Author Manuscript

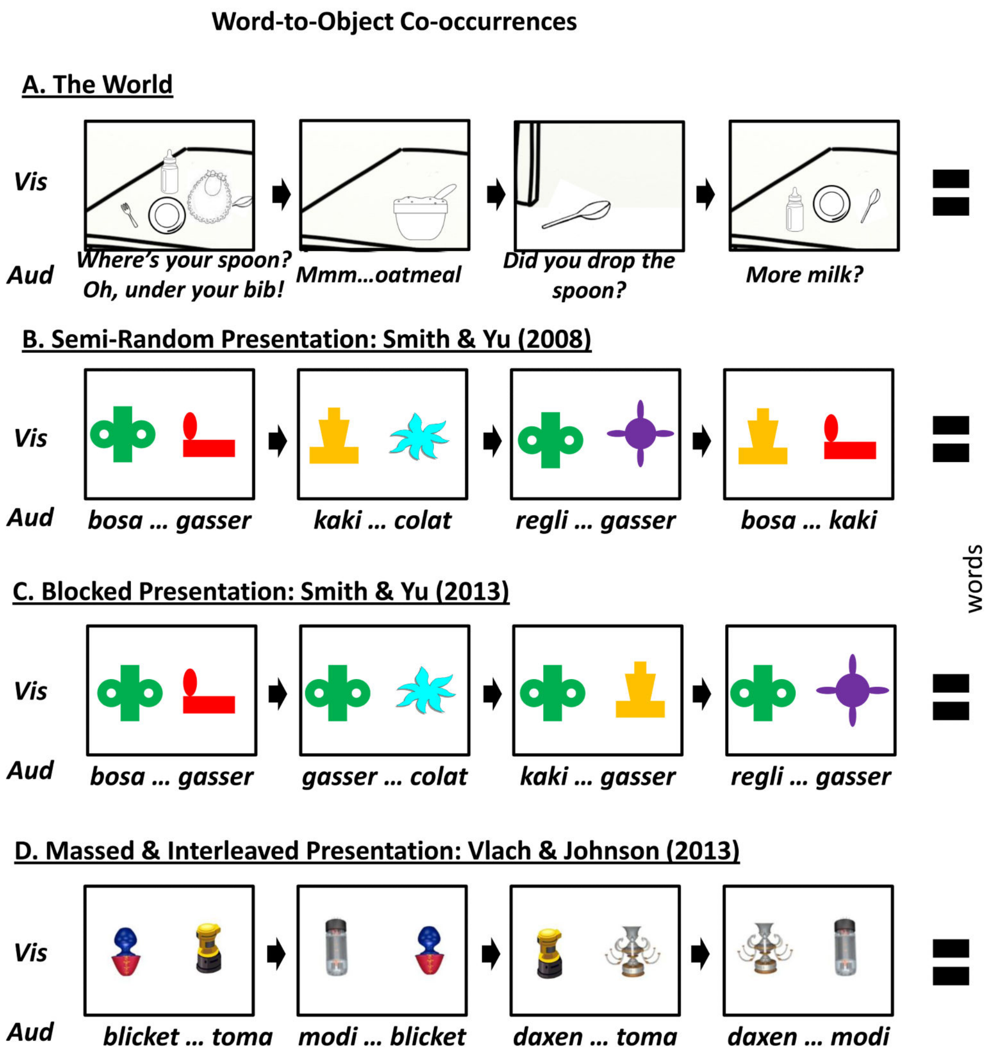
Highlights

Computational algorithms show that words-referent pairs be discovered in noisy data.

Infants use statistical regularities across trials to find word-referent pairs in noisy data.

Statistical learning offers a new solution to how infants break in to language.

The open question is the scalability of infant statistical learning to the real world.



Statistics

objects

words

	s	bi	o	m	p	f	bo	bt
spoon	2	1	1	1	1	1		1
bib	1	1		1	1	1		1
oat-meal	1		1				1	
milk	1			1	1		1	1

	b	r	k	g	c
bosa	2		1	1	
regli		1		1	
kaki	1		2		1
gasser	1	1		2	
colat			1		1

	b	r	k	g	c
bosa	1			1	
regli	1		1		
kaki		1	1		
gasser	1		1	3	1
colat				1	1

	b	y	g	w
blicket	2	1	1	
toma	1	2	1	
modi	1		2	1
daxen		1	1	2

Figure 1.

Series of scenes and words and the co-occurrence matrix. A. Example scenes and co-occurring language as be experienced by an infant. B. A sample of the scenes and co-occurring words in Smith & Yu [26]. Within a trial there was no information as to which object was the referent of each word. Across the 30 trials, each correct word-referent pair occurred 10 times and each spurious correspondence just twice. Trials (not shown) consisted of the presentation of two objects but one word. Looking duration to the target referent was the dependent measure. C. The "novelty trap" set by Smith & Yu [29] consisted of blocks of trials in which one object was repeated at the same location within a block and the other location showed nonrepeating objects. The final statistics were the same as in Smith & Yu [26] and testing was the same. D. The interleaved trials in the Vlach & Johnson [34] study, arranged so some word-referent pairs were adjacent and others were distant in the series. Final statistics and testing were comparable to the Smith & Yu studies.

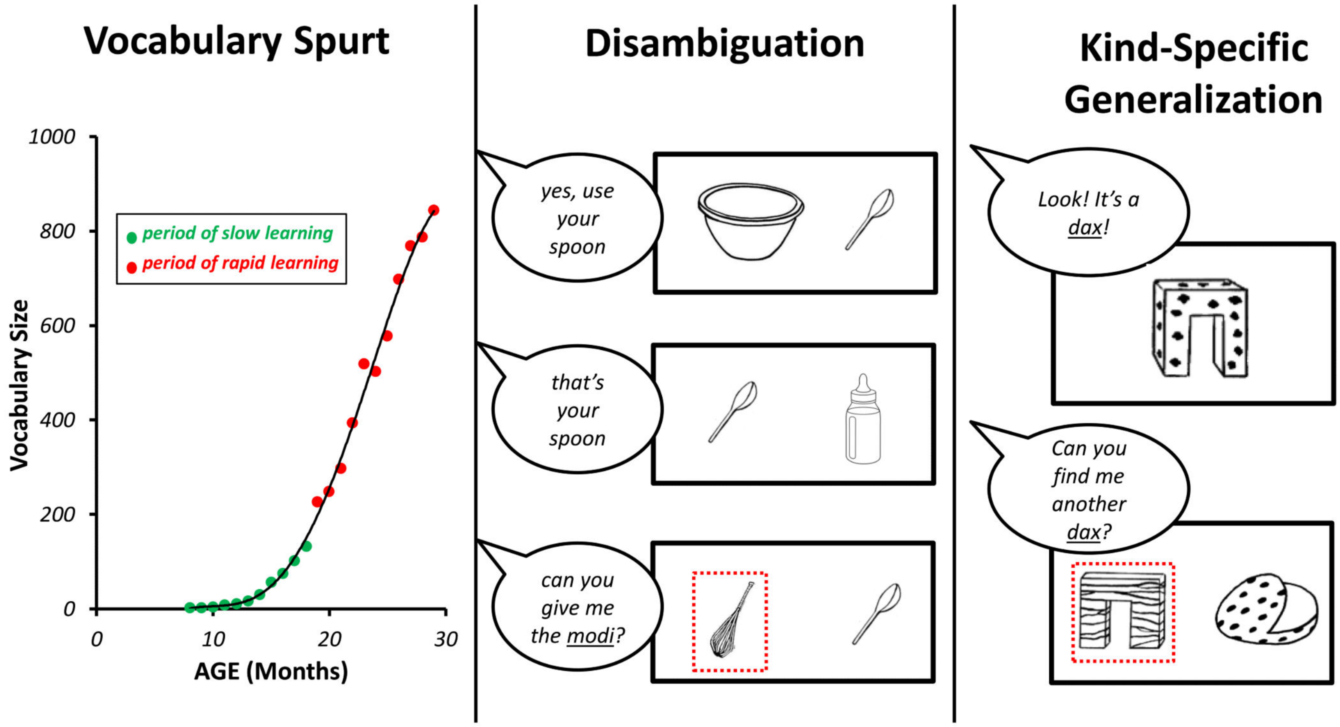
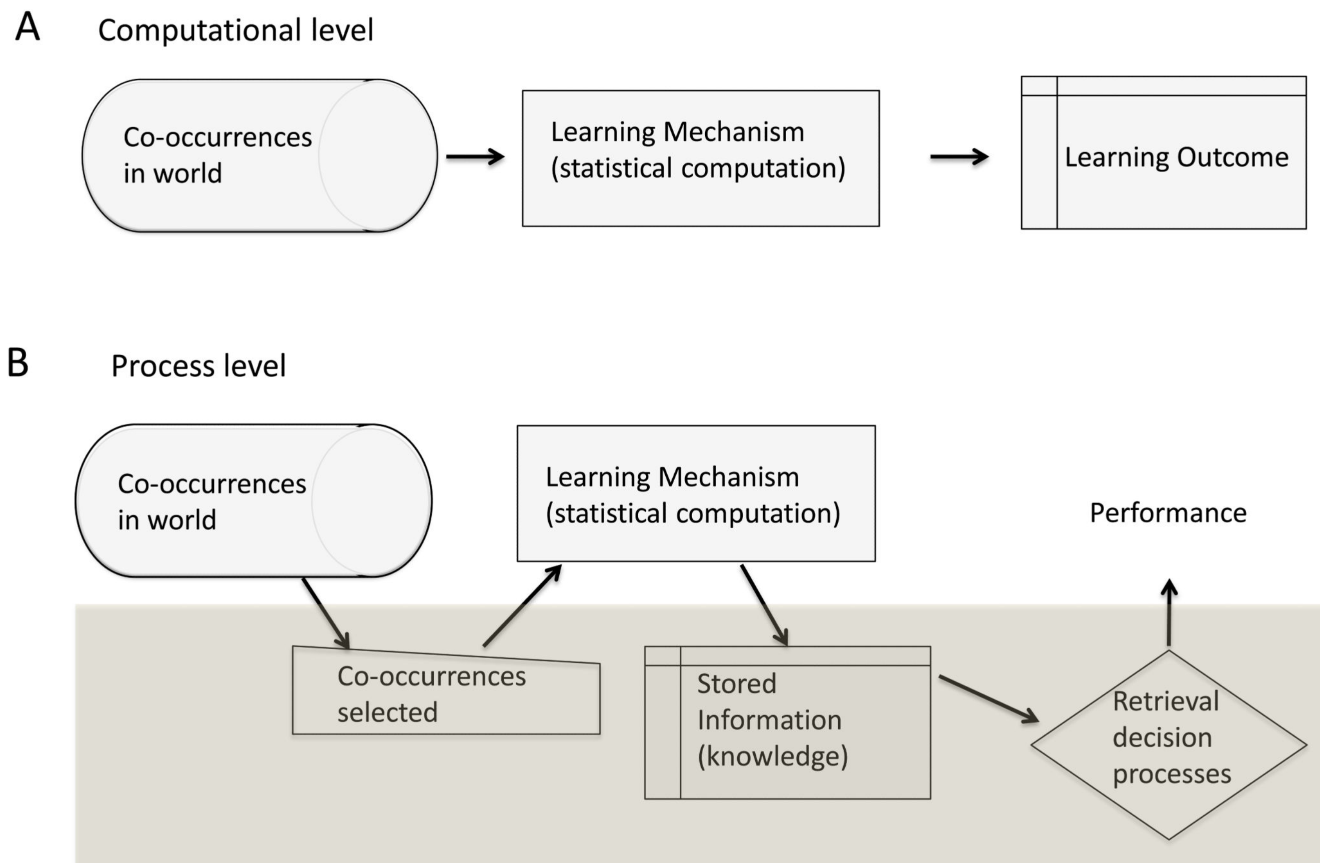


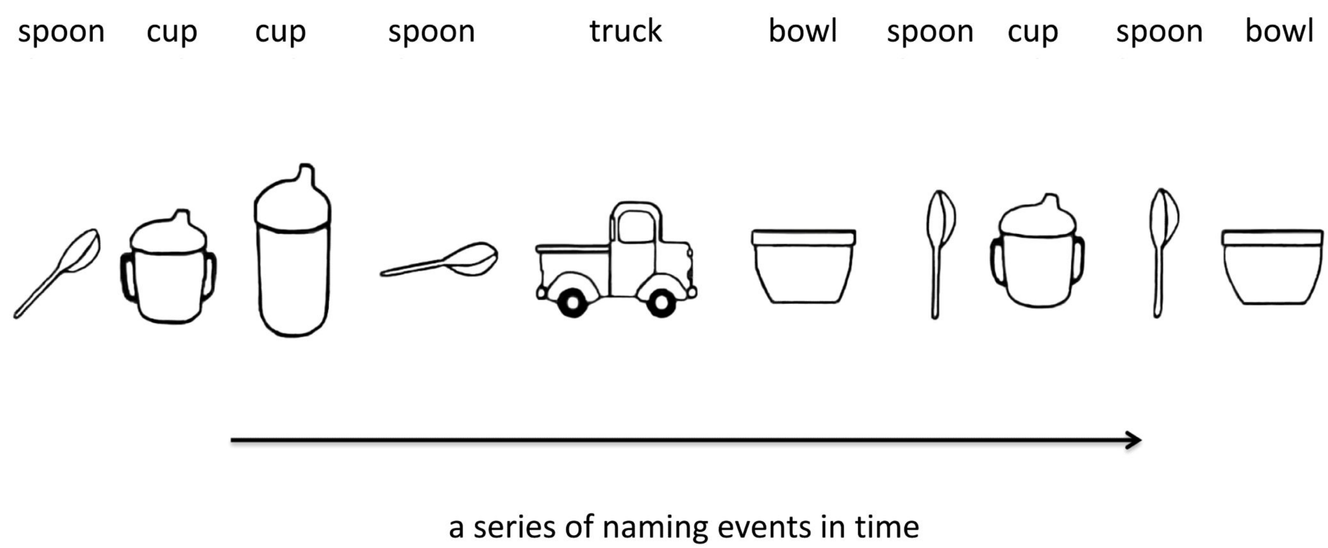
Figure 2.

Developmental changes in word learning. Children show knowledge of words and referents early, both in receptive and productive knowledge with new words added slowly at first and then more rapidly just prior to the second birthday [1-4,14, 15]. The period of rapid learning is also associated with increasingly sophisticated novel word learning strategies in measured in the laboratory including referential disambiguation [7, 15] and kind-specific generalizations [5, 18, 19]. In the disambiguation task children are reminded of the name(s) of an object(s) that they know and then a novel object is presented with a novel name and children consistently interpret the novel name as referring to the novel object. This disambiguation or fast mapping of a novel name to an object is sometimes considered an example of the mutual exclusivity principle of one name per object. The shape bias is an example of a kind-specific generalization. When children are told the name of a novel artifact, they generalize that name to other objects by similarity in shape; for substance-like novel objects, however, they systematically generalization the name by material [18].



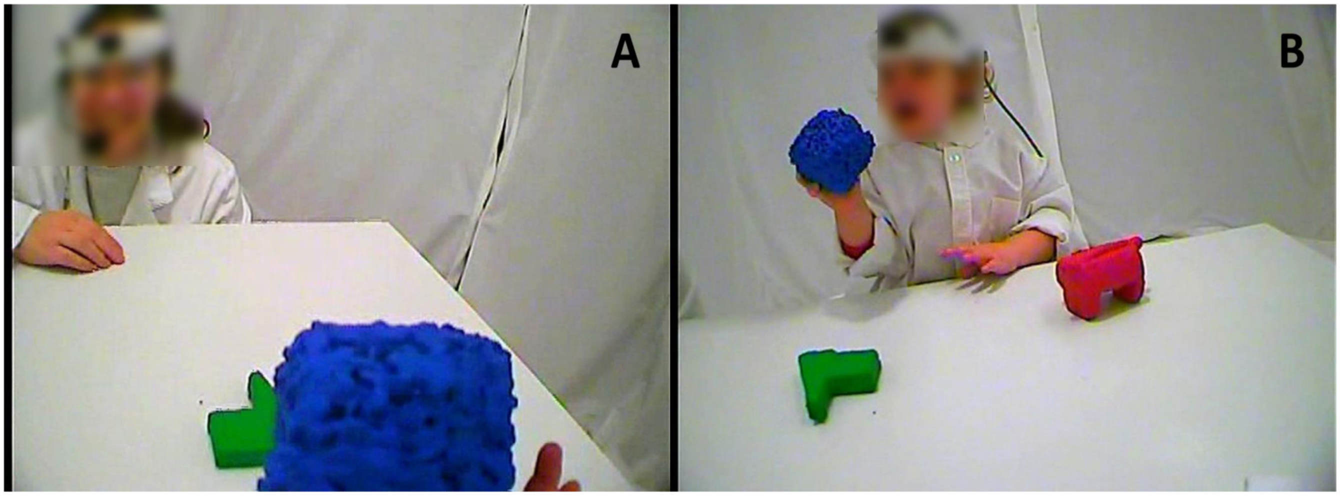
Box 1.

Many current models of cross-situational learning are formulated at a high conceptual level with just three components as shown in Panel A: the data (assumed or measured co-occurrences in the world), the learning mechanism (a set of statistical computations performed the data), and the demonstrated learning outcome. However, to simulate the performance of human learners, the model must –implicitly or explicitly -- make decision about other components of the learning system. Because these components interact in complex ways, these decisions can yield multiple but very different paths to the very same learning outcomes or performance.



Box 2.

A series of naming events and referents in time.



Box 3.

Image frames from an Infant-view head camera and from a Parent-view head camera as the parent and infant engage in joint toy play.