Statistical learning and language acquisition

Alexa R. Romberg* and Jenny R. Saffran

Human learners, including infants, are highly sensitive to structure in their environment. Statistical learning refers to the process of extracting this structure. A major question in language acquisition in the past few decades has been the extent to which infants use statistical learning mechanisms to acquire their native language. There have been many demonstrations showing infants' ability to extract structures in linguistic input, such as the transitional probability between adjacent elements. This paper reviews current research on how statistical learning contributes to language acquisition. Current research is extending the initial findings of infants' sensitivity to basic statistical information in many different directions, including investigating how infants represent regularities, learn about different levels of language, and integrate information across situations. These current directions emphasize studying statistical language learning in context: within language, within the infant learner, and within the environment as a whole. © 2010 John Wiley & Sons, Ltd. *WIREs Cogn Sci*

What is statistical learning? In its broadest sense, statistical learning entails the discovery of patterns in the input. This type of learning could range, in principle, from the supervised learning found in operant conditioning (learning that a certain behavior leads to reinforcement or punishment), to unsupervised pattern detection, to the sophisticated probability learning exemplified in Bayesian models. The types of patterns tracked by a statistical learning mechanism could be quite simple, such as a frequency count, or more complex, such as conditional probability. Likewise, the actual elements over which the computations are done could vary in complexity such as geometric shapes and faces, or in concreteness, such as syllables and syntactic categories.

The field of language acquisition has taken special interest in the idea of statistical learning because of the rapidity with which infants typically acquire their native language, despite the complexity of the structures to be acquired. The goal of this review is not to cover the well-trodden recent history of this area (for useful overviews, see Refs 1,2). Instead, we will highlight current directions in this field, with an eye toward the next phase of research on statistical language learning. A decade ago, the driving question in this area was whether infants actually track statistics in linguistic input. The answer to that question appears to be an unequivocal yes. Given that infants are clearly good pattern learners, the next set of questions concern how infants *use* those patterns.

This review is thus organized around some of the most interesting directions in which statistical language learning research is heading: *upward* through the levels of language structure beyond the initial task studied in this area, word segmentation; *inward* to connect with other cognitive mechanisms; and *outward* to ask whether statistics are actually useful given the rich input characteristic of natural languages. While this review will pose more questions than it will answer, we hope it will help to elucidate the next crucial steps for this burgeoning field of research.

In language acquisition, the term 'statistical learning' is most closely associated with tracking sequential statistics—typically, transitional probabilities (TPs)—in word segmentation or grammar learning tasks. A TP is the conditional probability of Y given X in the sequence XY. Typically, experimental materials are designed so that TPs can be calculated over the 'phonetic' content of the speech stream, such as segments, syllables, or words. However, a broad understanding of statistical learning incorporates both a greater range of possible computations and more

*Correspondence to: romberg@wisc.edu

Department of Psychology, University of Wisconsin, Madison, WI 53706, USA

aspects of the speech stream. It is possible that learners are computing any of several basic statistics such as frequency of individual elements, frequency of co-occurrence, mutual information, or many others. Prosodic patterns, stress patterns, distributional cues such as frequent frames, phonotactic patterns, the physical context of the interaction (e.g., objects in view), and the social context of the interaction (e.g., the speaker's eye gaze direction) could all enter into the computations of the learner. All of these types of regularities provide probabilistic information regarding language structure and use and are potentially helpful for learning about where words begin and end, lexical category membership, grammatical structure, and word meanings. While the primary focus of research to date has been demonstrating infant sensitivity to these regularities, it is also clear that no single cue is sufficient to acquire any aspect of language nor are cues independent of one another. The field is now moving toward an integrative approach: how do infant learners bring together multiple cues, both within domains (e.g., within the auditory stream) and across domains (e.g., between the auditory stream and the visual context) and examining how information is integrated and used over time (e.g., associating meanings with word forms that have been segmented using statistical cues).

UPWARD: APPLYING STATISTICAL LEARNING TO DIFFERENT LEVELS OF LANGUAGE

Studies of statistical language learning originated in questions concerning the sequential ordering of concrete elements, such as syllables.[3] While sequence learning is clearly of deep interest across many domains of knowledge, the field has expanded to examine potential statistical cues to linguistic structure across multiple levels of analysis, from phonology to grammar. Evidence is accumulating that statistical learning contributes to low-level processes like categorization of speech sounds, as well as higher level processes like word and grammar learning. These developments raise a number of interesting questions, including how learners 'know' which statistics to apply to which units of analysis, and how different levels of analysis interact with one another. For example, how does the output of one learning process become input to another learning process?

The question of how language statistics are represented and used for different levels of language is central to understanding how language acquisition proceeds during the first few years of life. In an influential series of studies, Jessica Maye and her colleagues examined how the acquisition of phonetic

categories is affected by the distribution of exemplars along an acoustic continuum.[4,5] Infants and adults exposed to a bimodal distribution of phonetic tokens are more likely to treat the distribution as consisting of two categories of elements than learners exposed to a unimodal distribution of the same elements. These findings suggest that learners group instances based on distributional as well as acoustic information, offering a clear example of how speech *perception* is shaped by the structure of the native language.

Distributional information can also reveal higher level structure. Words and phrases are initially opaque to learners; they are not clearly marked in the speech we hear. However, surface statistics signal the presence of these other levels of representation. Recent work offers evidence that infants are able to move from surface structure to deeper structure, such as tracking syllables to find words and then an underlying grammar[6] and tracking word-level computations to learn about phrasal units.[7–9]

Perhaps no place is this transition between levels more important than in discovering linguistic categories. In the absence of category structure, language users are limited to tracking the distributions of words. However, once learners discover the presence of categories, the nature of the learning problem changes from tracking statistics of observable tokens (words) to include information about more abstract types (linguistic categories). Corpus analyses suggest that distributional information should be highly relevant for category learning.[10] Surprisingly, research suggests that category learning via statistics, without other correlated cues, is challenging at best.[11–13] One exception is that adult learners use distributional information for categorization in the form of frequent frames[14]: words that consistently bound particular syntactic categories, such as '*you___it*' for verbs, or '*the___and*' for nouns.[15] These results were recently extended to include 12-month-old infants, who categorized novel words placed in highly familiar English frames.[16] It thus seems possible that distributional cues may powerfully facilitate categorization, particularly when combined with other phonological regularities that distinguish nouns and verbs.[17]

Statistical cues allow learners to do more than cluster elements together—they also allow learners to bridge levels of analysis. As learners track regularities in the speech stream, elements cohere in different ways, allowing the units over which computations are done to change with the learner's experience. In reality, this process probably involves complex interactions as different types of information become available and perceptual units or categories are refined and shaped. Consider studies of statistical

learning and word segmentation published over the past decade. Following exposure to fluent speech, successful discrimination between words and part-words (sequences spanning word boundaries) is taken as evidence that infants successfully segmented words. However, what discrimination *actually* demonstrates is that infants distinguish between sound sequences of varying internal coherence (e.g., high vs low TP). These results do not themselves speak to word *segmentation*. All that can be reasonably concluded is that infants have learned something about the statistics of the speech stream. While that is an important finding, it does not tell us whether statistical learning plays a role in the discovery of *words* in fluent speech. Note that this point applies generally to the broader infant segmentation literature, which relies on test discriminations of familiar versus novel words but has failed to investigate the representational status of those units.

To test word segmentation more directly, we developed a new task that combines methods from the word segmentation and word-learning literatures.[18] Seventeen-month-olds were first familiarized with a stream of continuous speech from a small artificial language, with only TP cues to word boundaries. After familiarization, the study diverged from the usual word segmentation task. Instead of testing infants on their ability to discriminate familiar and novel sequences (as measured by preferential looking), infants entered a label–object association task. Sequences from the word segmentation task were presented in isolation as labels for objects. Infants were habituated to the label–object pairs and then tested using the Switch procedure, designed for use in word-learning tasks.[19] On Same test trials, items consisted of labels and objects paired correctly, as observed during the habituation phase. On Switch test trials, the pairings were switched, violating the label–object associations presented during habituation. The logic behind this procedure is that if infants have learned the correct mappings and habituated to them, they should continue to be relatively uninterested in the Same trials, dishabituating only on the Switch trials (which contain incorrect pairings).

The critical manipulation concerned the status of the labels (see Ref 18 Exp. 2). For half of the infants, the labels presented during habituation were *words* from the speech stream heard during familiarization. For other infants, the labels were *part-words*—sound sequences spanning word boundaries. The words and part-words used as labels occurred equally often in the speech stream presented during familiarization. If statistical learning mechanisms generate representations based solely on familiarity with a string of sounds, the words and part-words should be equally good labels

for the novel objects. However, if statistical learning generated new representational units—candidate words, available for mapping to meaning—then the infants should more readily map word labels to meanings (here, objects) than the part-word labels. This is precisely what we found. Only infants for whom the labels were words showed a Switch effect on the test: longer looking times for Switch trials than Same trials. These results suggest that the statistics of the speech stream affected subsequent word learning, with infants more easily mapping statistically coherent sound sequences onto objects. Thus, infants did not only track statistics, but the output of the statistical learning process provided representations that served as good 'candidate words', available for mapping to meaning in the associative learning task, which involved tracking regularities between syllable sequences and an object presented visually. This is just one demonstration of how learning at one level of analysis could potentially affect learning downstream.[20–22]

INWARD: STATISTICAL LEARNING IN THE CONTEXT OF OTHER LEARNING MECHANISMS

While there is a consensus among researchers that statistical learning plays a role in language acquisition, the scope of this role is a hotly debated topic. It is one thing to show that infants behave in ways that demonstrate they are sensitive to the statistical structure of the input. However, this fact in and of itself does not illuminate the *process* of learning. And, as highlighted in the above discussion, few experiments have interrogated in detail the *nature of the representations* that are driving behavior on statistical learning tasks. Indeed, the term 'statistical learning' refers more to 'sensitivity to regularities in the input' than to a hypothesis about a particular mechanism of learning. Because of this lack of mechanistic understanding of statistical learning, it remains unclear how statistical learning is related to other types of learning hypothesized to play a role in language acquisition, including perceptual learning, hypothesis-testing, and rule learning.

It has turned out to be challenging to design experiments that clearly distinguish statistical learning-based accounts from rule learning-based accounts. In a paper that sparked much debate, Marcus and colleagues familiarized infants with strings of syllables that followed either an ABA or ABB pattern (e.g., *wo fe wo* or *wo fe fe*). Infants then discriminated strings of novel syllables that followed this familiarization pattern from those that did not.[23] The authors argued that because the test items had

no syllables in common with the familiarization items, TPs (or any other statistic) computed on the specific syllables presented during familiarization would not be informative during testing. Therefore, a statistical learning mechanism would not be sufficient to explain the infant's performance. They concluded that the infants employed a rule learning mechanism that operated over algebra-like variables. This interpretation has been challenged from two directions: (1) that statistical learning mechanisms are actually sufficient to explain the transfer[2,24–29] and (2) that repetition-detection is an automatic process of the auditory perceptual system.[30] There are a few different ways one could conceptualize transfer of an ABB pattern to novel strings within a statistical learning framework. It is possible that repetition is just another statistic that can be learned, such that infants are discriminating patterns of sames and differents (see discussions in Refs 23,25,31,32). Another perspective is that learning *during* the test session could account for the results, with the novel syllables being mapped onto the representations for the training syllables.[24,33] Under this view, a neural network would spontaneously learn to map the novel stimuli onto the internal representations learned during training. A third possibility is that prior learning specific to the speech stream (during word segmentation) created internal representations that allowed transfer to novel linguistic elements.[26] Importantly, each of these arguments regarding the flexibility of statistical learning entailed modeling the task in a neural network, rather than through further behavioral experiments. Each of these computational models relies on complex internal representations that are formed during performance of the task and drive the output of the model, sometimes in nonobvious ways. To the extent that these computational models are able to capture learners' behavior, they suggest that statistical learning is much more complex than simply tallying item-specific frequencies or conditional probabilities.

A separate but related challenge for statistical learning accounts of language acquisition is how infants know which regularities to track (or, under a multiple-learning-mechanism account, which learning mechanism to employ). One possibility is that the properties of the input itself determine how the input is processed. This hypothesis is currently being investigated in studies examining the circumstances under which learners can acquire nonadjacent dependencies—for example, the probability that A precedes B given an intervening X, as in AXB. Nonadjacent dependencies are difficult for even adult learners to acquire when they are presented in an artificial language with no other cues to grouping.[34,35]

However, when certain types of grouping cues are added, both adults and infants can successfully learn these structures. For example, Newport and colleagues found that when nonadjacent dependencies link speech sounds that shared acoustic features, such as consonants or vowels, adults were able to detect them[34] (see also Ref 36). High variability in intervening elements[37,38] also plays a role, though it is unclear whether variability provides a cue to grouping or causes the learner to shift from a default of tracking adjacent probabilities to looking for invariant structure in the midst of high variability. The ability to learn nonadjacent dependencies seems to develop during the second year of life, with a transition around 15 or 16 months—a finding that is supported by research using artificial language stimuli[38] and natural language stimuli.[39] However, recent work suggests that prior experience with adjacent dependencies can help even 12-month-old infants to detect related nonadjacent dependencies.[40]

Prior learning may also provide another type of grouping cue: familiarity with the elements in the input, and with their distributions, may make it easier to categorize elements of the input. Categorization could give learners easier access to less salient dependencies between the elements. Indeed, such a process may account for infants' success in discriminating the repetition grammars (ABA/ABB) discussed above. Infants are successful on this task when both training and test items are drawn from highly familiar categories such as speech sounds[23] and images of dogs or cats,[41] and when the items are multimodal.[42] Infants are also successful when the training set consists of speech sounds and the test set consists of other auditory stimuli, such as tones.[43] However, infants do not succeed at this task when the training set consists of auditory tones or a variety of other auditory or visual cues.[43] One interpretation of this set of results is that the familiarity of the elements in the training set (or perhaps the richness of the representation of those elements) influences the extent to which infants can generalize beyond the training set (see Ref 41, for discussion).

OUTWARD: RELATING STATISTICAL LEARNING TO REAL-WORLD LEARNING

The research to date clearly demonstrates that *in principle*, infants can track sequential statistics. However, these studies typically use artificial languages, presented either as synthesized speech streams or as natural speech lacking typical variability (e.g., syllables excised from monotone coarticulated speech and

recombined). Despite this artificiality, infants appear to process these materials as language, integrating them with native language information.[7,18,20,44–46] However, infants' ability to deploy statistical learning mechanisms given natural speech input remains unknown. While artificial languages afford researchers an unparalleled level of experimental control, the simplicity of these materials leads to concerns about ecological validity. For example, to eliminate cues other than particular regularity being tested, artificial materials typically use the same token of a particular syllable throughout the language (whether the token is synthesized or naturally produced). However, in natural language, the learner would need to determine that different tokens of a syllable represent examples the same type (i.e., that *dog* is a *dog* regardless of variability in pitch, intonation, or affect). Natural speech is exquisitely rich and complex and the learning mechanisms infants apply to a monotone, synthesized (or synthesized-sounding), pause-free, isochronous stream of speech may differ from those they apply to natural language 'in the wild'.

Alternatively, it is possible that the complexity of natural language actually facilitates learning. In particular, infant-directed speech contains attention-drawing prosodic manipulations, along with phonological cues that are often correlated with statistical cues. Even neonates prefer to listen to speech as compared to other environmental sounds.[47,48] And at least in artificial language studies, the presence of correlated cues typically facilitates learning.[12,22,49,50] However, it remains unclear how these learning mechanisms operate over natural speech. In one study using words marked with the correlated cues found in the Russian gender system, infants did successfully learn category structure.[51] However, no published studies have used natural fluent speech to assess statistical learning. It is possible that infants will fail when they are confronted with the variability inherent in natural speech (though see Ref 52, Exp. 11, for indirect evidence).

In a recent study, we combined the control of an artificial language with the variability of a natural language[53] in order to test infants' segmentation in a more ecologically valid context. The training corpus consisted of naturally produced Italian sentences. The target words were infrequent relative to previous statistical learning tasks and were surrounded by numerous other words, syllables, and phonemes. Infants discriminated test words with high TPs (the probability of *X* given *Y* in the sequence *XY)* from equally frequent words with low TPs. These results suggest that 8-month-olds can track statistical information across a corpus of naturally produced speech from a real language. A follow-up study demonstrated that 8-month-olds also track *backward* TPs presented in natural Italian speech.[54] These studies provide the beginnings of a research program in which specific statistical learning processes can be tested using realistic stimuli. In the absence of such studies, the relevance of statistical learning experiments to actual language acquisition will remain highly uncertain.

While there is much still to be learned about how infants track statistics in natural streams of speech, language learning does not happen in a sound-proof booth with nothing but an audio track. Rather, language learning takes place in context, with the infant and caregiver surrounded by objects they can see and touch and engaging in social interactions. The scope of studies of statistical learning of language has moved beyond the strict confines of speech itself to incorporate more of this rich context. Several recent papers have investigated how infants and adults might use cross-situational statistics to learn both the meanings of words[55–59] and the constraints that govern their acquisition.[60,61] For example, Smith and colleagues proposed that infant learners acquire a bias to extend object labels to similarly shaped objects by learning words for objects that come from categories that are well-defined by the physical shape of the members. On this view, the structure of infants' vocabularies leads infants to attend more readily to the shape of objects when learning new words.[60,61] Central to this account is the concept that the constraints that guide word learning are not independent of the input or the infant's experience. Instead, constraints emerge as infants learn about the ways that words are used and allocate attention to properties of objects that have been useful in the past.

Words are often used in ambiguous situations, in which there may be multiple possible referents present, leading to an inductive learning problem. Smith and Yu have suggested that one way to disambiguate word–referent pairings is to track the pairings over multiple scenes. For example, a learner might initially hear a label in the presence of object A and object B. In this case, it is unclear whether the referent of the label is object A or object B, leading to a failure to pinpoint the referent. However, if she subsequently hears the label in the presence of object B and object C, she might conclude that object B is the referent of the label, because while each instance is ambiguous in itself, object B consistently occurs with the label *across* instances. In fact, recent studies demonstrate that both adults[56,59] and 12- and 14-month-old infants[57] are able to capitalize on just such cross-situational statistics, learning multiple referent–label pairs in a short period of time by tracking pairs across a series of individually ambiguous situations.

While these cross-situational statistical computations are impressive, recent work suggests that they may be even more effective when a wider range of information is included. Social cues could be an important source of referential information. Frank and colleagues used a computational model to demonstrate that word meanings could be learned concurrently with learning about talker's referential intentions.[58] Their model uses a Bayesian framework and makes several predictions that are consistent with the constraints seen in word-learning tasks. Another computational model, using machine translation methods, explored how nonlinguistic cues could aid the learner in discovering how to map words to their real-world referents.[55] Indeed, the combination of joint attention, prosody, and co-occurrence statistics was more effective at learning word meanings than a model that used co-occurrence statistics alone. These studies show that language learning may be most efficient when regularities from the speech stream are combined with environmental regularities.

Another way to test the hypothesis that statistical learning is relevant to real language acquisition is to examine links between lab learning abilities and real-world language outcomes. This could be done via longitudinal designs, as others have done for studies examining other features of early language perception and processing.[62–64] In a recent study, we took a different approach: we tested a sample of grade-school aged children diagnosed with Specific Language Impairment (SLI) on a statistical learning task.[65] Compared with a group of typically developing children matched for age and nonverbal IQ, the children with SLI performed poorly on a task requiring tracking TPs in fluent speech from an artificial language. Strikingly, they also performed significantly worse than the comparison group on a nonlinguistic statistical learning task (tracking TPs of tones) with the same statistical structure as the language task. These results illuminate links between the lab learning abilities of these children and their native language outcomes. Moreover, the fact that the children with SLI underperformed on both the linguistic and nonlinguistic tasks suggests that the learning abilities linked to SLI are not limited to language (for related results with older children using a visual task, see Ref 66).

CONCLUSION

At this point, it is well established that infants are adept at tracking regularities in the speech stream. This review has focused on many of the directions that the field is now taking to study statistical language learning in a more complete context: within

language, within the infant learner, and within the environment as a whole. We end with some final comments regarding the major themes addressed by these divergent lines of research. Each of these themes highlights different ways that the field is moving from a very simple question 'Can infants track statistical dependencies in language?' to embracing the natural complexity of the language acquisition process. This move is imperative to a true understanding of language acquisition, as complexity is introduced from many different sources, including (though of course not limited to): the physical development of the infant learner, the rich hierarchical structure of language, the acoustic variability between talkers that the infant hears, and the many physical environments in which the infant experiences language and communicative acts. Ultimately, these sources of variability cannot be ignored, as we know that the process of language acquisition is likely to be more than the sum of its parts.

One of the most important themes to emerge from this body of work is the power of correlated cues. There are a number of ways in which cues could interact to aid language acquisition. Certainly, multiple cues could have an additive effect, such that learning is easier when more than one cue marks the structure to be learned. For example, children more easily learn how to generalize labels to different categories of items when the labels are presented in syntactic frames that reinforce the differences.[67] Correlated cues may serve to organize attention during learning, so that the learner can discover less salient structure. For example, nonadjacent dependencies and lexical categories are typically hard to learn from distributional information alone, but the presence of a correlated cue facilitates learning.[12,34,35,51] The correlation between cues may also lead to bootstrapping: using one cue allows for recognition of another cue that may eventually replace the first cue.[22] For example, in English, two-syllable words almost always follow a trochaic (strong–weak) stress pattern, and there is evidence that stress increasingly guides word segmentation during the first year of life.[46] However, infants cannot know the lexical stress pattern of their native language until they have successfully segmented some words. Infants are capable of tracking TPs from a very young age (see Ref 68 for data from neonates) and in linguistic and nonlinguistic domains.[69] A reasonable hypothesis is that young infants initially segment words using TPs, and as their lexicon develops, they are able to abstract the stress pattern in those words, allowing them to use stress in addition to, or in place of, TPs. Evidence for this hypothesis also comes from the finding that infants can abstract and generalize an

artificial phonological regularity (words begin with /t/) when it is consistent with TP information.[22]

The second major theme is the movement toward making statistical learning experiments more similar to real-world language learning, by using tasks that require generalization,[23,43,70,71] stimuli that are more similar to natural language[54] and tasks that move beyond discrimination and capture aspects of language use, such as mapping segmented words onto objects[18] and integrating across several instances or sources of information.[55–58] Studies concerning individual differences will also provide a powerful link between laboratory tasks and real-world outcomes.[65] These programs of research harness the control provided by laboratory tasks while admitting in the complexities of natural language structure and use. As the discipline continues to move upward, inward, and outward, we will be able to ask questions that scale up ever closer to the child's experience. Discovering the extent and limits of statistical learning abilities will help us to understand how children turn their linguistic experience into mastery of their native language.

REFERENCES

1. Gómez RL, Gerken L. Infant artificial language learning and language acquisition. *Trends Cogn Sci* 2000, 4:178–186.

2. Saffran JR, Werker JF, Werner LA. The infant's auditory world: Hearing, speech, and the beginnings of language. In: Kuhn D, Siegler RS, eds. *Handbook of Child Psychology: Cognition, Perception, and Language*. 6th ed. Vol 2. Hoboken, NJ: John Wiley & Sons; 2006, 58–108.

3. Goodsitt JV, Morgan JL, Kuhl PK. Perceptual strategies in prelingual speech segmentation. *J Child Lang* 1993, 20:229–252.

4. Maye J, Werker JF, Gerken L. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 2002, 82:B101–B111.

5. Maye J, Weiss DJ, Aslin RN. Statistical phonetic learning in infants: facilitation and feature generlization. *Dev Sci* 2008, 11:122–134.

6. Saffran JR, Wilson DP. From syllables to syntax: multilevel statistical learning by 12-month-old infants. *Infancy* 2003, 4:273–284.

7. Saffran JR. Words in a sea of sounds: the output of infant statistical learning. *Cognition* 2001, 81:149–169.

8. Saffran JR. Constraints on statistical language learning. *J Memory Lang* 2002, 47:172–196.

9. Saffran JR, Hauser MD, Seibel RL, Kapfhamer J, Tsao FM, et al. Grammatical pattern learning by infants and cotton-top tamarin monkeys. *Cognition* 2008, 107:479–500.

10. Mintz TH, Newport EL, Bever TG. The distributional structure of grammatical categories in speech to young children. *Cogn Sci* 2002, 26:393–424.

11. Frigo L, McDonald JL. Properties of phonological markers that affect the acquisition of gender-like subclasses. *J Memory Lang* 1998, 39:218–245.

12. Gómez RL, Lakusta L. A first step in form-based category abstraction by 12-month-old infants. *Dev Sci* 2004, 7:567–580.

13. Smith KH. Learning co-occurence restrictions: rule learning or rote learning? *J Verbal Behav* 1969, 8:319–321.

14. Mintz TH. Category induction from distributional cues in an artificial language. *Mem Cognit* 2002, 30:678–686.

15. Mintz TH. Frequent frames as a cue for grammatical categories in child directed speech. *Cognition* 2003, 90:91–117.

16. Mintz TH. Finding the verbs: Distributional cues to categories available to young learners. In: Hirsh-Pasek K, Golinkoff RM, eds. *Action Meets Word: How Children Learn Verbs*. New York: Oxford University Press; 2006, 31–63.

17. Monaghan P, Chater N, Christiansen M. The differential role of phonological and distributional cues in grammatical categorization. *Cognition* 2005, 96:143–182.

18. Graf Estes K, Evans JL, Alibali MW, Saffran JR. Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychol Sci* 2007, 18:254–260.

19. Werker JF, Cohen LB, Lloyd VL, Casasola M, Stager CL. Acquisition of word-object associations by 14-month-old infants. *Dev Psychol* 1998, 34:1289–1309.

20. Thiessen ED, Saffran JR. Learning to learn: infants' acquisition of stress-based strategies for word segmentation. *Lang Learn Dev* 2007, 3:73–100.

21. Lany J, Saffran JR. From statistics to meanings: Infant acquisition of lexical categories. *Psych Sci* 2010, 21:284–291.

22. Sahni SD, Seidenberg MS, Saffran JR. *Connecting cues: Overlapping regularities support cue discovery in infancy*. Manuscript in press, Child Development (In press).

23. Marcus GF, Vijayan S, Bandi Rao S, Vishton PM. Rule learning by seven-month-old infants. *Science* 1999, 283:77–80.

24. Altmann GTM, Dienes Z. Rule learning by seven-month-old infants and neural networks. *Science* 1999, 284:875a.

25. Seidenberg MS, Elman J. Do infants learn grammar with algebra or statistics? *Science* 1999, 284:434–435 (author reply 436-7).

26. Christiansen MH, Curtin S. Transfer of learning: rule acquisition or statistical learning? *Trends Cogn Sci* 1999, 3:289–290.

27. Marcus GF. Reply to Seidenberg and Elman. *Trends Cogn Sci* 1999, 3:289–289.

28. Marcus GF. Reply to Christiansen and Curtin. *Trends Cogn Sci* 1999, 3:290–291.

29. Marcus GF. 'Reply to Christiansen and Curtin': Corrigendum. *Trends Cogn Sci* 1999, 3:322–322.

30. Gervain J, Macagno F, Cogoi S, Peña M, Mehler J. The neonate brain detects speech structure. *Proc Natl Acad Sci U S A* 2008, 105:14222–14227.

31. McClelland JL, Plaut DC. Does generalization in infant learning implicate abstract algebra-like rules? *Trends Cogn Sci* 1999, 3:166.

32. Seidenberg MS, Elman JL. Networks are not hidden rules. *Trends Cogn Sci* 1999, 3:288–289.

33. Dienes Z, Altmann GTM, Gao S-J. Mapping across domains without feedback: a neural network model of transfer of implicit knowledge. *Cogn Sci* 1999, 23:53.

34. Newport EL, Aslin RN. Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognit Psychol* 2004, 48:127–162.

35. Peña M, Bonatti LL, Nespor M, Mehler J. Signal-driven computations in speech processing. *Science* 2002, 298:604.

36. Perruchet P, Tyler MD, Galland N, Peereman R. Learning nonadjacent dependencies: no need for algebraic-like computations. *J Exp Psychol Gen* 2004, 133:573–583.

37. Gómez RL. Variability and detection of invariant structure. *Psychol Sci* 2002, 13:431–436.

38. Gómez R, Maye J. The developmental trajectory of nonadjacent dependency learning. *Infancy* 2005, 7: 183–206.

39. Santelmann LM, Jusczyk PW. Sensitivity to discontinuous dependencies in language learners: evidence for limitations in processing space. *Cognition* 1998, 69:105–134.

40. Lany J, Gómez RL. Twelve-month-old infants benefit from prior experience in statistical learning. *Psychol Sci* 2008, 19:1247–1252.

41. Saffran JR, Pollak SD, Seibel RL, Shkolnik A. Dog is a dog is a dog: infant rule learning is not specific to language. *Cognition* 2007, 105:669–680.

42. Frank MC, Slemmer JA, Marcus GF, Johnson SP. Information from multiple modalities help 5-month-olds learn abstract rules. *Dev Sci* 2009, 12:504–509.

43. Marcus GF, Fernandes KJ, Johnson SP. Infant rule learning facilitated by speech. *Psychol Sci* 2007, 18: 387–391.

44. Chambers KE, Onishi KH, Wu Y, Lomibao J. Statistical learning and word recognition: Nonwords and words mingle. *Bienniel Meeting of the Society for Research in Child Development*, Boston, MA, 2007.

45. Swingley D, Aslin RN. Lexical competition in young children's word learning. *Cognit Psychol* 2007, 54:99–132.

46. Thiessen ED, Saffran JR. When cues collide: use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Dev Psychol* 2003, 39:706–716.

47. Vouloumanos A, Werker JF. Tuned to the signal: the privileged status of speech for young infants. *Dev Sci* 2004, 7:270–276.

48. Vouloumanos A, Werker JF. Listening to language at birth: evidence for a bias for speech in neonates. *Dev Sci* 2007, 10:159–164.

49. Morgan JL, Meier RP, Newport EL. Structural packaging in the input to language learning: contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognit Psychol* 1987, 19:498–550.

50. Thiessen ED, Hill EA, Saffran JR. Infant-directed speech facilitates word segmentation. *Infancy* 2005, 7:53–71.

51. Gerken L, Wilson R, Lewis W. Infants can use distributional cues to form syntactic categories. *J Child Lang* 2005, 32:249–268.

52. Jusczyk PW, Houston DM, Newsome M. The beginnings of word segmentation in English-learning infants. *Cognit Psychol* 1999, 39:159–207.

53. Pelucchi B, Hay JF, Saffran JR. Statistical learning in a natural language by 8-month-old infants. *Child Dev* 2009, 80:674–685.

54. Pelucchi B, Hay JF, Saffran JR. Learning in reverse: 8-month-old infants track backward transitional probabilities. *Cognition* 2009, 113:244–247.

55. Yu C, Ballard DH. A unified model of early word learning: integrating statistical and social cues. *Neurocomputing* 2007, 70:2149–2165.

56. Yu C, Smith LB. Rapid word learning under uncertainty via cross-situational statistics. *Psychol Sci* 2007, 18:414–420.

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

58. Frank MC, Goodman ND, Tenenbaum JB. Using speakers' referential intentions to model early cross-situational word learning. *Psychol Sci* 2009, 20:578–585.

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

60. Smith LB, Jones SS, Landau B, Gershkoff-Stowe L, Samuelson L. Object name learning provides on-the-job training for attention. *Psychol Sci* 2002, 13:13–19.

61. Smith L, Samuelson L. An attention learning account of the shape bias: reply to Cimpian & Markman (2005) and Booth, Waxman & Huang (2005). *Dev Psychol* 2006, 42:1339–1343.

62. Newman R, Ratner NB, Jusczyk AM, Jusczyk PW, Dow KA. Infants' early ability to segment the conversational speech signal predicts later language development: a retrospective analysis. *Dev Psychol* 2006, 42:643–655.

63. Marchman VA, Fernald A. Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Dev Sci* 2008, 11:F9–F16.

64. Tsao F-M, Liu H-M, Kuhl PK. Speech perception in infancy predicts language development in the second year of life: a longitudinal study. *Child Dev* 2004, 75:1067–1084.

65. Evans J, Saffran JR, Robe K. Statistical learning in children with specific language impairments. *J Speech Lang Hear Res* 2009, 52:321–335.

66. Tomblin JB, Mainela-Arnold E, Zhang X. Procedural learning in adolescents with and without specific language impairment. *Lang Learn Dev* 2007, 3:269–294.

67. Yoshida H, Smith LB. Linguistic cues enhance the learning of perceptual cues. *Psychol Sci* 2005, 16:90–95.

68. Teinonen T, Fellman V, Näätänen R, Alku P, Huotilainen M. Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neurosci* 2009, 10:21.

69. Kirkham NZ, Slemmer JA, Johnson SP. Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition* 2002, 83:B35–B42.

70. Gerken L. Nine-month-olds extract structural principles required for natural language. *Cognition* 2004, 93:B89–B96.

71. Gómez RL, Gerken L. Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition* 1999, 70:109–135.