CHAPTER

13

Statistical Learning Mechanisms in Infancy

J. Lany[1], J.R. Saffran[2]

[1]University of Notre Dame, Notre Dame, IN, USA; [2]University of Wisconsin-Madison, Madison, WI, USA

OUTLINE

While psychologists long imagined that infants experience the world as a bewildering array of sights, sounds, and sensations (James, 1890), research since the 1950s has shown that infants are born with more advanced perceptual and cognitive abilities than was once thought. For example, infants' visual, auditory, tactile, and vestibular systems are functional at birth, and their discrimination abilities in these areas are already remarkably acute (see Kellmann and Arterberry, 2000 for a comprehensive discussion). They show evidence of processing and remembering stimuli in their environment; even very young infants differentially attend to repeated stimuli versus novel stimuli (e.g., Fagan, 1970). Infants also exhibit differential attention to stimulus features. For example, newborns prefer complex, patterned stimuli to homogeneous ones (Olson and Sherman, 1983), facelike stimuli to nonfacelike ones (Valenza

et al., 1996), and speech to other acoustic stimuli (Vouloumanos et al., 2010).

In addition to advances in our understanding of infants' perceptual abilities, there has been great progress in the last two decades in the study of the mechanisms by which infants learn about their environment. Scientists began studying whether infants can learn via classical and operant conditioning, and by the 1970s, researchers had made substantial progress in discovering the conditions under which such learning occurred (Rovee-Collier, 1986; Sameroff, 1971). In the decades since, researchers have continued to make exciting discoveries about the nature of the learning mechanisms by which infants become sensitive to structure in their environment. In this chapter, the focus is on recent research on infants' ability to learn four different kinds of statistical structure: probability distributions, sequential

structure, correlations between stimulus dimensions, and associations between different forms of information. These types of statistical structures appear to be foundational to many aspects of development, including auditory and visual perception, language development, early cognition, and event processing. Much of our discussion is focused on the role of these learning mechanisms in language acquisition, though connections are also made between language learning and learning in other domains. Whenever possible, research investigating the relationships between behavioral findings and underlying neural processes has also been presented. The studies reviewed in this chapter reveal the power, nuance, and complexity of infant statistical learning mechanisms and contribute to an understanding of their role in early cognitive development.

13.1　LEARNING PROBABILITY DISTRIBUTIONS

13.1.1　Tracking Probability Distributions in Speech Sounds

A basic form of learning available to infants is sensitivity to the frequencies with which events occur in the environment. Early research on infant perceptual development revealed that infants can differentiate between frequent and infrequent events: when presented with successive trials of two pictures side by side in which one is repeated while the other changes, infants spend more time looking toward the novel picture (Fantz, 1964). Recent studies revealed that by the time they are about 8 months old, infants can do something potentially much more powerful than tracking frequencies of individual events: they can track the relative frequencies of items within a set, or probability distributions.

Differences between speech sounds, or phonemes, that correspond to differences in the meanings of words are called 'phonetic contrasts.' For example, the English sounds /r/ and /l/ are considered to be different phonemes, representing a phonetic contrast, because a switch from one to the other changes the meaning of a word (i.e., 'right' vs. 'light'). However, while /r/ and /l/ are perceived differently by adult speakers of English, they are not by speakers of Japanese, as these speech sounds do not correspond to a meaningful difference in their language. Thus, adults' ability to discriminate among speech sounds reflects the phonetic organization of their native language; adults are best able to discriminate between sounds that embody phonetic contrasts.

While adults' phonetic perception differs according to their native language, landmark research in the 1970s revealed that early in development, infants across languages readily discriminate among almost all of the speech sounds of the world (e.g., Eimas et al., 1971; for review, see Aslin et al., 1998a). Within the first year of life, however, infants' phonetic discrimination patterns become closely matched to the phonetic contrasts relevant in their native language (Kuhl et al., 1992; Werker and Tees, 1984). Within a given language, the distribution of speech sounds along a particular continuum appears to reflect phonetic category information (e.g., Lisker and Abramson, 1964). Specifically, every production of a speech sound differs on a variety of acoustic dimensions, influenced by the characteristics of individual speakers' vocal tract, speech rate, coarticulation, and phrase or sentence-level prosody. However, for acoustic dimensions that are critical in the perceived differences between phonemes, the values will tend to form nonuniform distributions, in which some values along a dimension are much more frequent than others (i.e., forming a bimodal or trimodal distribution). Interestingly, the differences relevant for highlighting phonemic contrasts may be exaggerated in infant-directed speech (Burnham et al., 2002; Kuhl et al., 1997; Werker et al., 2007). Adults can capitalize on nonuniform variation in speech sounds to discriminate between phonemes (Kluender et al., 1998), and computational modeling studies suggest that sensitivity to probability distributions may play an important role in learning phonetic contrasts (McMurray et al., 2009; Valhaba et al., 2007).

Maye et al. (2002) tested whether experience with unevenly distributed sound profiles influences phonetic perception in English learning in infants at 6 and 8 months of age. They created a set of eight speech sounds forming a continuum between [da] and [ta]. The endpoints of the continuum differed in the amount of voicing in the initial aspiration, from [da] (voiced) to [ta] (unvoiced), with six intermediate values. Infants in the unimodal condition heard a distribution in which the intermediate values (tokens 4 and 5) occurred most frequently, with decreasing frequencies toward the tails of the distribution. For infants in the bimodal condition, tokens 2 and 7 occurred with high frequency, while the tokens at the midpoint and endpoints were relatively less frequent.

After the infants were familiarized to the unimodal or bimodal distribution, they were tested to determine whether they discriminated between sounds at the opposite endpoints of the [da] – [ta] continuum, using a preferential listening task. In this procedure, the presentation of an auditory stimulus is contingent on infants' attention to a visual stimulus, such as a flashing light or a checkerboard. As long as the infants continue to look toward the visual stimulus, the auditory stimulus continues to play, but once they look away, it stops and a new trial begins. If infants' phonetic discrimination is affected by the distributional properties of the speech

sounds, they should continue to discriminate between stimuli at the endpoints when the distributional information supports the maintenance of two phonetic categories. However, when the values of the tokens form a unimodal distribution, infants should treat the sounds as belonging to a single phonetic category, ignoring variation in voicing. Only infants in the bimodal condition showed discrimination of the endpoints, suggesting that experience with unimodal distributions along an acoustic continuum may play a role in the loss of sensitivity to phonetic contrasts not relevant in their language.

Recent research suggests that experience with distributional variation in speech input can also result in an enhancement or sharpening of discrimination. For example, Kuhl et al. (2006) found that at 6–8 months of age, both English- and Japanese-learning infants discriminate between /r/ and /l/, a phonetic contrast present only in English. By 10–12 months of age, the Japanese-learning infants showed reduced discrimination, consistent with other studies showing a loss of discrimination of nonnative language contrasts at this age. the English-learning infants, on the other hand, showed better discrimination between /r/ and /l/ at 10–12 months than at 6–8 months. Moreover, infants who hear speech in which acoustic differences in phonemic contrasts are exaggerated also show enhanced speech discrimination skills (Liu et al., 2003), suggesting that they capitalize on acoustic information differentiating speech sounds.

Maye et al. (2008) tested the role of distributional learning in the enhancement of speech-sound discrimination using a contrast that is difficult for infants to discriminate: the contrast between pre-voiced and short-lag stop consonants, such as [da] versus [ka] (Aslin et al., 1981). They found that 8-month-old infants learning English failed to discriminate the endpoints when given no additional exposure to speech sounds on the continuum, confirming that this is a difficult distinction for infants. Infants exposed to a unimodal distribution of tokens in the lab also failed to show discrimination. However, infants exposed to a bimodal distribution showed evidence of discrimination, suggesting that sensitivity to distributional information can lead to enhanced discrimination of difficult speech-sound contrasts. Interestingly, at 10–12 months of age, just after the period of perceptual reorganization, infants' speech-sound discrimination abilities show diminishing effects of experience with distributional information. Specifically, they need more extensive experience with a bimodal distribution to show evidence of discrimination of speech sounds that are not contrastive in their native language (Yoshida et al., 2010).

In sum, developmental changes in infants' phonetic perception appear to result, at least in part, from their ability to track distributional regularities in their native language. Kuhl (2004) has proposed that infants' experience with acoustic and distributional cues works in concert with maturational changes in neural development to produce such changes. Data from neurophysiological recordings suggest that there are neural changes in infants' processing of nonnative phonetic contrasts. Cheour et al. (1998) used an oddball paradigm to test Finnish infants' discrimination of native and nonnative vowel contrasts at both 6 and 12 months of age. In these studies, a vowel present in the infants' native language was repeated, with another 'oddball' vowel presented on about 1 in 10 trials. The oddball was either another native-language vowel, which differed in the frequency of the second formant, or a vowel that does not occur in Finnish. Evoked-response potential (ERP) studies using the oddball paradigm have shown that when infants detect a change to the repeating stimulus, the ERP wave shows a negative deflection beginning at about 200 ms, called a mismatch negativity, or MMN. At 6 months of age, Finnish infants showed a similar MMN to both the native and nonnative oddball vowels, but by 12 months, they showed a much greater MMN response to the native vowel. Using a similar paradigm in a longitudinal study with English-learning infants, Rivera-Gaxiola et al. (2005b) found an enhancement in the neural response to native-language contrasts between 7 and 11 months. They also found that infants continued to show a neural response indicative of discrimination for nonnative vowels, though the neural manifestation of the discrimination took two different forms. In one group of infants, the oddball elicited a negative component, similar to their response at 6 months of age, while in a second group, it elicited a greater positive component. Interestingly, Rivera-Gaxiola et al. (2005a) found that infants who continued to show a negative component in response to nonnative oddball phonemes later had smaller native-language vocabularies, while those that showed a positive response had larger vocabularies. The authors hypothesized that the development of the positive response in infants who later developed more advanced language skills reflects neural reorganization in response to native-language patterns. These data are in line with Kuhl's (2004) hypothesis that changes in neural organization may play a role in changes in phoneme perception during infancy.

In sum, experience with distributional information plays an important role in the dramatic changes in speech-sound discrimination that take place between 8 and 10 months of age. While infants remain sensitive to distributional cues marking nonnative phonemic contrasts after the period of reorganization, these findings suggest that neural changes, increased experience with the distributional properties of one's native language, or a combination of these factors, reduces their impact on discrimination. It is important to note that the social

context in which infants experience nonnative language distributional patterns may also modulate the effects of such patterns, but by increasing sensitivity to novel phonetic contrasts – infants show greater sensitivity to difference in nonnative contrasts after listening to a native speaker in person than after watching and listening to a native speaker on a video tape (Kuhl et al., 2003). These findings suggest that although reorganization takes place early and has dramatic effects on speech perception, the capacity to learn novel contrasts is maintained and can be facilitated through social interaction.

13.1.2 Tracking Probabilities in Sampling Events

The studies described so far suggest that distributional information may facilitate perceptual reorganization in the auditory realm over the first year. Another line of recent research suggests that sensitivity to the distributional regularities plays a pervasive role in learning and cognition across diverse areas of development. Inductive learning involves encountering a finite set of examples (a sample), and using that information to extrapolate beyond the sample to behave flexibly when encountering new examples. To do so, learners must use information about the distribution of samples to learn about the underlying properties of the population from which the samples were drawn.

In a recent series of studies, Xu and Garcia (2008) explicitly tested infants' sensitivity to the relationship between the distribution of individual samples and the underlying populations they belong to. In one experiment, 8-month-old infants watched an experimenter bring out a covered box and remove 5 balls of mostly one color, either 4 red and 1 white, or 4 white and 1 red ball with her eyes closed. After she had removed the 5 balls, the experimenter removed the box's cover. Thus, the infants could not see the composition of balls in the box until after the sample balls were removed. The infants looked at the box longer on trials when the sample was unlikely under conditions of random sampling (i.e., when the container contained mostly red balls after the experimenter had just drawn mostly white balls from it). This suggests that infants' knowledge of the properties of the sample influenced their expectations about the properties of the population (i.e., the covered box). Interestingly, when the same experiment was repeated, but the experimenter drew the ping-pong balls from her pocket, infants showed no differences in looking times toward boxes containing mostly red versus mostly white balls. In other words, when the test balls were not sampled from the box, infants did not respond differentially to the contents of the box as a function of the properties of the sample. This suggests that in the

initial experiment, infants were responding to the relation between a sample and the population they were drawn from, rather than to more superficial perceptual relationships between the sample balls and the balls in the box.

Teglas et al. (2007) also investigated whether infants' processing of sampling events is affected by population information. Twelve-month-olds viewed a movie depicting a container with a small opening at its base, with four objects bouncing inside. Critically, three of the objects were identical, and one was unique. Next, an occluder covered the container, and one of the objects exited the container through the opening. On half of the trials, the infants saw events with probable outcomes (one of the three identical objects exited), and on the other trials, they saw the less probable outcome (the unique object exited the container). The infants looked longer at the improbable than the probable outcome, suggesting that the composition of objects in the container influenced their processing of the outcomes. However, when the container was partitioned such that only the unique object was physically capable of exiting, the infants looked longer when one of the three identical objects exited the container. This suggests that they were not simply responding to surface information about object frequency, looking longer toward a less common or less familiar object, but rather were responding to the relationship between the population of objects in the container and the outcome of the sampling event.

These studies also hint that infants are sensitive to information about the sampling process itself. For example, when the sampling process is random, as when an experimenter with her eyes closed picked balls out of a box in Xu and Garcia (2008), infants expect the sample to reflect the global properties of the population. However, when the sampling was not random, infants showed evidence of adjusting their expectations about the relations between a sample and the population. For example, when an occluder made some sampling events impossible in Teglas et al. (2007), infants did not appear to expect that the more frequent elements would be sampled. Moreover, Xu and Denison (2009) found that infants use information about an individual's preferences (e.g., for red balls) to guide their sampling behavior, and they are not surprised when she looks into a box containing mostly white balls and picks out mostly red ones.

Taken together, these findings suggest that across domains, infants are highly sensitive to the probabilities of individual events and can integrate this information in order to track the probabilities of sets of items related to one another in an underlying distribution or population. In the domain of speech perception, this process appears to facilitate differentiation between random noise and meaningful variations in similar sounds. In the

domain of event perception, sensitivity to distributional information allows infants to generalize from individual experiences to anticipate future scenarios.

13.2 LEARNING CO-OCCURRENCE STATISTICS

Within a given domain, the reliable co-occurrence of elements is typically a surface manifestation of a meaningful relationship between those elements. Co-occurrence information thus has the potential to play a pervasive role in learning, as long as learners are sensitive to it. Despite the seemingly simple nature of such associative relationships, tracking sequential associations in most environmental patterns can be quite challenging. It entails encoding a stimulus, as well as those that precede and follow it, and maintaining memory representations across multiple occurrences of each stimulus. This might not seem difficult to do for one item that occurs in a reliable context, but consider tracking this information for a set of 100, or even 1000 different items that occur in highly variable contexts.

Natural language provides a compelling illustration of this difficult problem. Given that the sounds comprising words and sentences unfold over time, as opposed to being expressed simultaneously, spoken language is rich with sequential structure. Language is hierarchically organized, consisting of patterns at both very fine-grained and larger-grained levels, and there are sequential regularities at each level of structure. In part because of the high demands that tracking such complex sequential structure would place on infants, the potential role of co-occurrence learning mechanisms in language acquisition has been the object of a great deal of interest. Numerous studies have tested two related hypotheses: (1) tracking sequential associations could lead to the discovery of language structure, both within words as well as across words and in simple grammatical patterns, and (2) infants have sufficient computational resources to track such complex information.

13.2.1 Sequential Learning: Phonotactics

At a very fine-grained level, languages are organized according to phonotactic patterns: regularities in how sounds are structured within words. What makes these regularities interesting is that they differ across languages, and thus must be learned. For example, English syllables can begin with some consonant clusters (e.g., /st/ and /fr/), but not others (e.g., /fs/ and /ng/). However, none of these sequences can appear in syllable onsets in Japanese. Infants appear to be sensitive to phonotactic patterns by 9 months of age (Friederici and Wessels, 1993; Jusczyk et al., 1993, 1994). In addition, 9-month-old infants can learn novel phonotactic patterns given brief laboratory experience with novel languages (e.g., Saffran and Thiessen, 2003). Following exposure to lists of novel words in which the initial consonant within a syllable was always voiced (e.g., /g/, /d/, /b/) and the final consonant was unvoiced (e.g., /k/, /t/, /p/), infants were able to distinguish between novel words consistent with those patterns and ones that followed the opposite pattern. However, when voicing was not a consistent feature of syllable onsets and offsets (i.e., /g/, /b/, and /t/ occur in syllable-initial position, while /k/, /d/, and /p/ occur in syllable-final position), infants failed to show evidence of learning the phonotactic regularities. This suggests that infants can learn positional regularities when they follow a consistent pattern, but that learning restrictions on the positions of individual elements is taxing for infants at this age.

Interestingly, by 16.5 months, infants successfully learn individual restrictions on sound locations within syllables (Chambers et al., 2003). These results suggest that with age and linguistic experience, infants' ability to learn more specific regularities is enhanced. Because learning an abstract pattern typically results from the ability to track regularities or similarities across individual exemplars, it is often more demanding than tracking information about specific exemplars. However, when exemplars share a salient feature, the need for tracking exemplar-specific information can be reduced and may lessen the computational demands on learning. This may be why younger infants were able to track phonological regularities only when they conformed to a phonological generalization.

13.2.2 Sequential Structure: Word Segmentation

Another source of sequential regularity in language pertains to how syllables are combined to form words. Syllables that reliably co-occur often belong to the same word, while syllables that rarely co-occur are more likely to span word boundaries (e.g., Swingley, 2005). Because of this feature of language, transitional probabilities between syllables tend to be higher within words than between words. The transitional probability (TP) of a co-occurrence relationship between two elements, X and Y, is computed by dividing the frequency of XY by the frequency of X. This yields the probability that if X occurs, Y will also occur. Saffran et al. (1996) tested whether 8-month-old infants, who are just beginning to learn their first words, can capitalize on TP cues. Infants listened to a stream of synthesized speech in which the only potential cue to word boundaries was the TPs between adjacent syllables (1.0 within words; 0.33 at word boundaries). The infants listened to the

stream for about 2 min and were then tested on their ability to discriminate between the syllable sequences with high TPs ('words') and novel combinations of the familiar syllables ('nonwords'). Infants listened longer to the strings containing nonwords, suggesting they distinguished between syllable sequences that were attested in their input (TPs = 1.0) and sequences that did not occur (TPs = 0). A second group of infants was familiarized in the same manner and tested on sequences with high TPs 'words' (all internal TPs were 1.0), and 'part-words': sequences of syllables that had occurred across word boundaries (e.g., the final syllable of one word and the first two syllables of another word), which had lower TPs of 0.33. Infants again displayed a novelty preference, suggesting that they can also discriminate sequences that contain highly reliable transitions from those that contain less reliable ones.

Subsequent studies more specifically probed the kinds of sequential statistics that infants are sensitive to in these studies, in particular, distinguishing between the frequency of a sequence and its transitional probability. The frequency of a sequence is simply the number of times it occurs, while the transitional probability of a co-occurrence relationship provides a measure of how tightly linked or connected X and Y are, controlling for the raw frequency of X. If X occurs many times without Y, then no matter how many times it occurs with Y, the conditional probability will be relatively low. Conversely, a sequence can have low frequency but a high conditional probability.

In the Saffran et al. (1996) study, the syllable transitions within words were both more frequent and had higher TPs than the syllable transitions in part-words. Aslin et al. (1998b) thus investigated whether infants can track TPs, or just frequencies. Specifically, they modified the design of Saffran et al. (1996) such that two of the words occurred twice as often as the other words. This manipulation permitted a design in which the four test items – two words and two part-words – were equally frequent; however, the TPs within words were 1.0, and the TPs spanning the word boundaries were still 0.33. Infants showed discrimination between the words and part-words, despite the fact that the syllable sequences occurred equally often in both types of test items. This suggests that by 8 months of age, infants have a powerful mechanism for tracking co-occurrence relationships and for distinguishing potentially spurious co-occurrences from ones in which there is a very tight connection.

Another recent study further specified the nature of the regularities that infants can use to learn reliable co-occurrence relationships. The TPs in these studies have primarily been described as prospective relationships, or the probability that, given the occurrence of a syllable, another syllable will follow (i.e., the probability that the

syllable 'by' will follow the syllable 'ba,' as in the word baby). However, infants could also be sensitive to the probabilities of retrospective relationships (i.e., the probability that the syllable 'ba' will precede the syllable 'by'). In the studies of sensitivity to TPs reviewed so far, the words contained forward TPs (FTPs) and backward TPs (BTPs) of 1.0, and thus infants could have used either or both to discriminate words from part-words. Moreover, in natural languages, FTPs and BTPs are both likely to be reliable cues to word boundaries, and thus tracking both would potentially be advantageous to infants. Pelucchi et al. (2009a) thus familiarized 8-month-old infants with a corpus in which HTP and LTP words differed in their BTPs, but both had FTPs of 1.0. Thus, infants could only use BTPs to distinguish between the HTP and LTP words. Infants showed significant discrimination between HTP and LTP words, suggesting that they were sensitive to the BTPs of the syllable sequences. While FTPs may facilitate anticipating upcoming stimuli, infants' sensitivity to BTPs may help them to remember what came before a particular element. Thus, infants' sensitivity to both types of relationships potentially enhances word segmentation and other tasks that require sequential learning and processing.

In most of the studies described so far, infants were given limited exposure to artificial languages, produced as synthetic speech, which is not as engaging as naturally spoken infant-directed speech. More recent studies have investigated whether infants also use these cues to segment speech when there are many more competing regularities to attend to, and many more individual segments over which to compute TPs. For example, Pelucchi et al. (2009b) tested whether infants can track forward TPs in a natural language. They presented English-learning 8-month-olds with Italian sentences spoken with infant-directed prosody. Embedded in these sentences were two words with high internal TPs (1.0) and two words with low internal TPs (0.33). Over the course of the 2-min familiarization phase, infants heard each of these words presented just 18 times. When tested, infants discriminated between high TP words and low TP words, suggesting that they were able to track TPs despite the complexity and variability of the materials. Pelucchi et al. (2009a), described above, also used naturally spoken Italian in their work showing that infants can track backward TPs.

In another study examining how statistical learning mechanisms scale up to the complexity of natural language, Johnson and Tyler (2010) tested infants' ability to use TPs in a corpus in which the words varied in syllable number. They familiarized 8- and 5.5-month-old infants to a synthesized speech stream in which two bisyllabic words and two trisyllabic words were concatenated in a continuous stream. Interestingly, infants failed to discriminate words from part-words

under these conditions. This suggests that infants benefit from the presence of other regularities beyond TPs (i.e., consistent syllable length). Indeed, learners also appear to benefit from rhythmic patterns that point to word boundaries, though the degree to which this occurs depends on the infants' age, presumably reflecting the amount of native-language knowledge the infant has acquired (e.g., Johnson and Jusczyk, 2001; Theissen and Saffran, 2003, 2007).

Altogether, these results provide strong evidence that infants are sensitive to sequential relationships between phonemes and syllables. The extent to which such cues interact with other potential cues to word boundaries, such as stress patterns, sentential prosody, and the occurrence of words in isolation, are areas particularly in need of further study.

13.2.3 Sequential Learning: Grammatical Patterns

At yet another level up in the hierarchy of language structure, sensitivity to sequential relationships appears to play a role in learning grammatical patterns, such as learning how words can be combined into phrases and sentences. For example, Saffran and Wilson (2003) asked whether sensitivity to word boundaries arising from tracking TPs between syllables provides a foundation for tracking word combinations. They played 12-month-old infants synthesized speech strings in which the TPs between syllables served as a reliable cue to word boundaries (i.e., TPs within words were 1.0, while TPs of syllables spanning word boundaries were 0.25). The strings also contained word-order patterns that were not directly cued by relationships between syllables, but could only be detected by tracking the TPs between the words. Infants were then tested on their ability to discriminate novel grammatical strings from ungrammatical ones. Despite the fact that the TPs of adjacent syllables could not be used to distinguish between the grammatical and ungrammatical test strings, infants showed significant discrimination. These findings suggest that sensitivity to sequential relationships resulting in segmentation may bolster subsequent learning by helping infants to learn sequential relationships between words.

Gómez and Gerken (1999) also tested whether 12-month-old infants can learn TPs within multiword 'sentences.' They created an artificial language in which nonsense words were combined to form sentences in which there were probabilistic regularities in the ordering of words. Infants were then tested on novel grammatical and ungrammatical strings that contained familiar words. Infants successfully discriminated between grammatical and ungrammatical strings,

providing evidence that they learned the probabilistic co-occurrence relationships between words. In a subsequent experiment, infants were also able to distinguish between grammatical and ungrammatical strings when the strings were instantiated in a novel vocabulary. This finding suggests that learning sequential statistics can lead not just to knowledge of individual sequences but also potentially to a more abstract level of structure.

13.2.4 Learning Nonadjacent Co-occurrence Probabilities

The studies on sequential learning described thus far suggest that infants readily track predictive relationships between adjacent segments such as syllables and words. However, in natural languages, predictive relationships can also occur between nonadjacent elements. For example, grammatical dependencies marking tense are often nonadjacent, as in the relationship between auxiliaries such as '*is*' and the progressive inflection '*ing*,' as they are necessarily separated by a verb (e.g., '*is running*,' '*is eating*,' '*is talking*'). Likewise, a plural noun predicts plural marking on the subsequent verb, but the noun and verb can be separated by modifiers, as in '*The kids who were late to school are in trouble*.' Tracking nonadjacent dependencies places greater demands on memory than tracking adjacent dependencies, as elements must be remembered long enough to be linked to other elements occurring later in time. In addition, because nonadjacent dependencies can be separated by several word elements, there are many potentially irrelevant relationships for the learner to track, presenting a considerable computational burden.

Given the demands involved in detecting nonadjacent regularities, it is perhaps unsurprising that both infants and adults have substantial difficulty learning them. For example, while infants can track the relationships between adjacent elements well before they turn a year old, infants start showing sensitivity to grammatical relationships involving nonadjacent elements in their native language only at about 18 months of age (Santelmann and Jusczyk, 1998), suggesting that some combination of language experience and maturation of neural substrates for memory is needed to facilitate nonadjacent dependency learning.

Gomez (2002) investigated the conditions that promote sensitivity to nonadjacent relationships. In particular, she hypothesized that there is a relationship between the presence of salient adjacent structure and the tendency to track nonadjacent structure. Because tracking adjacent relationships appears to be relatively easy for infants, they are likely to focus on adjacent structure as long as it is reliable. However, when the variability in adjacent relationships is high (i.e., when adjacent TPs are low),

infants may be less likely to track those relationships and more likely to track reliable nonadjacent regularities.

To test this hypothesis, Gomez (2002) exposed 18-month-olds to artificial language strings containing non-adjacent dependencies with TPs of 1.0. Critically, these dependencies were separated by an intervening element drawn from a set of 3, 12, or 24 different elements, and thus the adjacent TPs between words varied across conditions: relatively high (0.33, set size 3), medium (0.08, set size 12), or very low (less than 0.01, set size 24). Only infants in the set size 24 condition, in which the predictability of adjacent sequences was very low, successfully discriminated between familiar grammatical strings and strings that violated the nonadjacent relationships.

In subsequent research, Gómez and Maye (2005) found that 15-month-old infants track nonadjacent dependencies even under conditions of high variability but that 12-month-olds fail to do so. This developmental pattern suggests that increases in memory capacity over the second year facilitate nonadjacent dependency tracking. In addition to developments in memory, another factor in the development of nonadjacent dependency learning is prior language experience. Just as segmentation facilitates learning higher-order patterns in which those elements are combined, Lany and Gómez (2008) found that prior learning about adjacent relationships facilitates infants' detection of nonadjacent patterns. In particular, when 12-month-old infants were first given experience with adjacent relationships between word categories, they then successfully detected novel nonadjacent relationships between words from those categories.

13.2.5 Learning Co-occurrence Statistics in Other Domains

The studies described so far suggest that infants' sensitivity to sequential regularities could play a role in detecting multiple layers of language structure, from phonotactics to grammar. Far from being specific to learning in the auditory domain, infants' sensitivity to sequential structure plays an important role in learning across domains. Sequential relationships are important dimensions of event structure, such as in brushing one's teeth (e.g., picking up the toothbrush, applying toothpaste, wetting the brush, and brushing), making coffee, or opening a toy container. The subcomponents of the action tend to co-occur in reliable sequences, and thus tracking sequential structure in the visual domain may play a role in both detecting event boundaries and in learning their internal structure.

Indeed, Baldwin et al. (2008) found that adults can use statistical information to learn regularities in dynamic action sequences. Employing a design similar to studies investigating word segmentation (e.g., Saffran et al., 1996), adults viewed a series of discrete actions spliced together. There were action triplets embedded within the stream such that within a sequence of individual actions (e.g., poke, scrub, drink), TPs were 1.0, and across action sequences, the TPs between actions were 0.33. The only cues to the action sequence boundaries were the TPs. Adults were able to discriminate trained action sequences (TPs of 1.0) from unattested sequences of familiar actions (TPs of 0.0), as well as from sequences spanning boundaries (TPs of 0.33). They also showed evidence of detecting sequential regularities in action sequences when the individual actions (e.g., the action of drinking) varied considerably in their perceptual features across training.

While similar studies with action sequences have not yet been done with infants, numerous studies suggest that infants track sequential statistics in other domains of visual processing. For example, Kirkham et al. (2002) tested infants' ability to detect sequential relationships in a series of shapes. In these experiments, 2-, 5-, and 8-month-old infants were habituated to a series of six shapes that appeared sequentially on a screen. The objects were grouped into three sets of sequential pairs. Within a pair, one shape was reliably followed by another shape; the TPs between objects within a pair were 1.0, while the TPs at pair boundaries were 0.33. At all of the ages tested, infants discriminated between strings that preserved the statistical regularities and sequences in which the sequential regularities were violated. While infants in this experiment could use either frequency of co-occurrence or TPs to distinguish the test trials, Marcovitch and Lewkowicz (2009) found that by 4 months of age, infants can separately track frequency of shape co-occurrence as well as the conditional probability of shape pairs. These findings suggest that the ability to track sequential co-occurrence relationships in the visual domain emerges quite early in development.

Other studies have focused on infants' ability to learn sequential regularities in dynamic events. Kirkham et al. (2007) showed 8- and 11-month-old infants a spatiotemporal sequence in which a red dot appeared consecutively in the six positions of a 2×3 grid. There were reliable 'location pairs' in the sequence: for example, if the dot first appeared in the top left, it always subsequently appeared in the top middle. After the second element in a pair, the dot could appear in one of three locations. Thus, within the continuous sequence, the TPs within a location pair were 1.0, and TPs spanning pairs were 0.33. Only the 11-month-olds distinguished between sequences in which location pairs were preserved and sequences containing the unattested transitions. However, when each location in the habituation sequence was occupied by a different object, 8-month-olds

distinguished between test sequences in which the location pairings were preserved and those in which they were violated. Eye tracking data revealed that infants were also faster to reach the second element in a location pair, which was predictable given the previous element, than to initial elements, which were not predictable given the last object/location, providing additional evidence that they had learned the spatiotemporal regularities. These results suggest that learning location-based co-occurrence relationships is challenging, developing between 8 and 11 months of age, despite robust abilities to track purely temporal co-occurrence relationships that develop many months earlier. However, infants can succeed at younger ages when the sequential relationships are also marked by featural cues, or when the occurrence of the object at one location predicts not just a subsequent location but also the identity of the object that will appear next.

13.2.6 Neural Correlates of Learning Sequential Structure

Given the behavioral results amassed over the past decade, researchers have become extremely interested in the neurophysiological processes involved in tracking sequential regularities. Cunillera et al. (2006, 2009) recorded ERPs as adults listened to a continuous stream of 3-syllable words cued by TPs, similar to the materials used by Saffran et al. (1996). They observed an increased negativity (an N400) occurring just after the onset of syllables in word-initial position appearing over the course of familiarization with the stream. Abla et al. (2008) found a similar pattern when learners were presented with continuous streams of 'tone words,' an effect that was more pronounced in participants who showed better discrimination in a subsequent forced-choice test. This is similar to behavioral studies by Saffran et al. (1999), which showed successful segmentation of tone sequences using TPs. Sanders et al. (2002) measured ERPs to trisyllabic words presented in isolation before and after hearing those words presented in continuous strings with TPs as the only cues to word boundary locations. Despite the difference in their methods, they also found that words evoked a greater N400 after training, as well as an N100 at syllable onset. Moreover, ERP recordings of sleeping newborn infants who were played similar streams of continuous syllables show a negative deflection of the ERP wave during the first syllable of a word (Teinonen et al., 2009). These findings suggest that there is a reliable neural signature related to tracking TPs in the auditory domain.

Interestingly, similar ERP signatures have been observed in adults learning sequential co-occurrence relationships in *visually* presented stimuli. For example, Abla and Okanoya (2009) found that triplet onsets within

a visually presented stream of shapes also evoked an N400. In other words, just as in the case of syllable and tone streams, the first element of a statistically defined unit within a stream of shapes evoked a distinctive response. There was also overlap in the areas in which the increased N400 was observed across these studies – specifically middle frontal and central sites – across both auditory and visual modalities.

Together, the findings suggest that the N400 reflects something about learning reliable sequences across domains. In many studies of language processing, N400 components reflect the occurrence of an unexpected word, and thus its appearance at the first syllable of statistically defined words could reflect the fact that its occurrence was less predictable than the internal syllables. It could also reflect a search for the representation of the sequence as the first syllable is heard.

13.3 LEARNING CORRELATIONAL STRUCTURE

Another form of learning involves tracking correlations among properties of objects and events. Similar to tracking a sequence of events such as syllables or objects, tracking correlational structure entails learning associations, and thus there is potential overlap in the learning mechanisms involved. In the case of learning correlational structure, however, the relevant associations lie between features or dimensions of an object or event (such as an object or its label) rather than sequential components as they unfold over time. First, recent research suggesting that tracking correlations facilitates word segmentation and learning grammatical categories is described. Research suggesting that sensitivity to correlational structure plays an important role in segmenting the visual world, as well as in forming object categories (such as learning to distinguish between cats and dogs or between plates and spoons) is also discussed.

13.3.1 Sensitivity to Correlated Cues in Language Acquisition

In the previous section, evidence that tracking sequential structure facilitates finding word boundaries in fluent speech (Aslin et al., 1998a,b; Saffran et al., 1996), as well as some evidence suggesting that sensitivity to sequential TPs may interact with other phonological regularities in word forms was reviewed. Specifically, Johnson and Tyler (2010) found that when word length is consistent, even 5.5-month-old infants can track TPs in continuous speech. However, when word length varies, infants have a harder time using TPs to segment the stream. This suggests that infants' ability to use TPs as

a segmentation cue may be facilitated by the presence of correlated cues such as regular word length.

Sahni et al. (2010) investigated whether sensitivity to TPs marking word boundaries can facilitate learning other correlated cues relevant to segmentation. They exposed 9-month-old infants to a continuous stream of bisyllabic words in which TPs within words were 1.0 and TPs across word boundaries were 0.25. Critically, all the 'words' began with /t/ (e.g., *tohsigh*, *teemay*, *tiepu*). The question of interest was whether infants could use TP information to learn enough about the speech stream to discover the /t/-initial cue. Infants successfully discriminated between novel words that had /t/ onsets and words in which /t/ occupied a medial location, suggesting they had learned that words begin with /t/. Critically, infants failed this same discrimination test when they had no prior familiarization to the speech stream, or when they were familiarized with a stream in which every other syllable began with /t/ but there were no TPs indicating word boundaries. These results suggest that infants can learn correlations between word properties to aid segmentation – in this case, using a known cue (TPs) to discover a novel correlated cue to word boundaries (/t/-onsets).

Evidence that infants can learn sequential relationships between words by 12 months of age has already been reviewed (Gómez and Gerken, 1999; Saffran and Wilson, 2003). However, a critical property of grammatical structure is that it pertains not just to how individual words are combined (e.g., 'the cat' is grammatical, but 'cat the' is not) but also to how words from different grammatical categories can be combined (e.g., determiners like 'the' and 'a' precede nouns rather than follow them). While tracking distributional information is useful for learning word-order patterns, learning grammatical categories with distributional cues alone can be very difficult (Braine, 1987; Smith, 1969). Importantly, in natural languages, words from different syntactic categories are distinguished both by distributional (Cartwright and Brent, 1997; Mintz, 2003; Mintz et al., 2002; Monaghan et al., 2005; Redington et al., 1998) and phonological cues (Farmer et al., 2006; Kelly, 1992; Monaghan et al., 2005). In other words, words from the same grammatical category share correlated distributional and phonological features. For example, nouns tend to occur after both 'a' and 'the,' and to have a strong–weak stress pattern. Critically, Braine (1987) found that the presence of correlated cues facilitates learning the category-level co-occurrence relationships in adults. Braine hypothesized that the presence of the additional cues reduces the computational demands involved in tracking a large number of individual co-occurrence relationships between specific words. Rather than tracking and accurately remembering each individual co-occurrence relationship, the presence of correlated cues may allow learners to detect the associations between the phonological features shared by words within a category.

To test whether infants can also use correlated cues to form grammatical categories, Gerken et al. (2005) exposed 17-month-old English-learning infants to Russian words drawn from two grammatical categories. In Russian, words have complex morphological structure, often consisting of a stem plus multiple grammatical morphemes. The familiarization set consisted of six feminine and six masculine words, and all of the words contained additional grammatical morphemes: feminine words ended in the case markers 'oj' and 'u' and masculine words ended in the case markings 'ya' and 'em.' The case markings provided distributional cues to the feminine and masculine categories. An additional phonological cue marking the category distinction was present on half of the words: three of the feminine words contained a derivational suffix 'k,' and three of the masculine words contained the suffix 'tel.' Thus, in many feminine words, 'k' was followed by 'oj' and 'us' (e.g., 'polku' and 'polkoj'), while in many masculine words, 'tel' was followed by 'ya' and 'em' (e.g., 'zhitelya' and 'zhitelyem'). Infants familiarized with these words were subsequently able to distinguish between novel grammatical words containing those relationships and ungrammatical ones: for example, even if they had not heard 'zhitelyem,' they were able to distinguish it from the ungrammatical 'zhitelu.' They were also able to distinguish between 'vannoj' (grammatical combination) and 'vannya' or 'vannyem' (ungrammatical combinations). In this case, they could not have been using the co-occurrence relationships between the case markers and derivational morphemes (the 'telya' and 'koj' sequences), but were generalizing based on distributional information.

Thus, when word categories are marked by correlated cues sharing phonological properties and distributional characteristics, infants successfully learn and generalize the category relationships. While 17-month-olds show evidence of generalizing based on distributional information, Gerken et al. (2005) found that 12-month-olds failed to do so. However, using a similar category-learning paradigm, Gómez and Lakusta (2004) found that 12-month-olds can learn correlations between distributional and phonological features. These findings are consistent with the hypothesis that category learning may initially involve tracking the correlations between distributional and phonological features of words.

13.3.2 Learning Correlational Structure in the Visual Domain

Infants' visual environments are exceptionally complex. Nevertheless, elements that reliably co-occur across time and space tend to belong to the same object

and thus provide information about object boundaries. Thus, just as in the case of segmenting words from continuous speech, the co-occurrence of features in the visual field is potentially a powerful cue that could be used to learn what clusters form objects. To test whether infants can track such conditional probabilities in complex visual displays, Fiser and Aslin (2002) showed 9-month-old infants scenes containing three elements. In each scene, three discrete shapes were presented simultaneously in a 2×2 grid. Two of the elements formed a 'base pair': those objects always occurred together in a consistent spatial orientation (the probability of co-occurrence, or TP, was 1.0). The two elements of each base pair also occurred with a third element, but its position relative to the base pair varied (TP = 0.25). Infants were then tested on scenes containing two elements – either base pair elements in their proper configuration or two elements that did not form a base pair. Infants were able to discriminate between base pairs and nonbase pairs when they differed in the frequency with which they appeared during habituation, and also when they were equally frequent but differed in their TPs. This suggests that sensitivity to statistical information facilitates object perception and raises the question of whether such learning can support object perception in younger infants.

Category learning is another domain in which sensitivity to correlations between object attributes might facilitate learning. Object categories (e.g., cups, dogs, birds, and trees) are structured such that properties characteristic of the category tend to co-occur within individual instances of the category. For example, the presence of one feature of a tree (branches, leaves, bark, trunk) is correlated with the presence of the others, but less strongly correlated with the presence of features associated with objects from other categories (e.g., ceramic handles, fur, and beaks).

Younger (1985) asked whether infants' sensitivity to such correlations plays a role in category learning. Ten-month-old infants viewed drawings of a set of animals that were composed of several continuously varying features (i.e., leg length, tail width, neck length, and ear separation). In the 'broad' condition, infants were exposed to a set of animals with features that were uniformly distributed (e.g., animals with long necks had both long and short legs), forming one category with a broad distribution of feature values. In the 'narrow' condition, the feature values formed two correlated clusters: for example, long-necked animals all had short legs, while short-necked animals had long legs. In this condition, the animals clustered together into two potential categories, each with a narrow range of features.

Infants were habituated to a set of these animals and then tested with novel animals. One animal contained feature values that were the average of all the animals in the familiarization set for the broad condition, but falling in between the average of the two narrow categories. The second animal contained a set of features that were closer to the average of one of the categories in the narrow condition, but farther from the average of the broad category. Infants in the narrow condition dishabituated to the broad stimulus, while infants in the broad condition dishabituated to the test animal closer to the average of one of the narrow categories. This suggests that when feature values formed two clusters, infants formed categories corresponding to correlations between values on these dimensions. However, when feature values were randomly distributed, infants did not group the animals into separate categories.

In similar studies, Younger and Cohen (1986) found that 7-month-olds, but not 4-month-olds, are sensitive to correlations among object features. However, Mareschal et al. (2005) noted that because 4-month-old infants appear to have learned some perceptually defined categories, for example, the difference between squares and diamonds (Bomba and Siqueland, 1983), and cats and dogs (Eimas and Quinn, 1994), the studies of Younger and colleagues may have underestimated infants' categorization abilities. They hypothesized that for young infants with limited memory, longer exposures to individual animals may make it more difficult to keep all the animals in memory. In particular, if infants can only retain information about one animal at a time, they would be unable to detect the correlations between object features across a set of animals. Thus, they modified the habituation phase such that the individual animals were presented for briefer durations. Under these conditions, 4-month-olds showed evidence of forming categories based on correlations between features.

These findings suggest that tracking correlated cues plays an important role in learning perceptually based object categories. There is also a potentially important connection between infants' ability to track correlations between features and their sensitivity to distributional information. Specifically, studies of distributional learning, such as Maye et al. (2002, 2008), suggest that infants can use the frequencies across a single dimension of a complex stimulus to form groupings. In these studies, infants show evidence of tracking correlations among the values of several distributions.

Quinn et al. (2006) began to investigate the neural underpinnings of categorization in infants. In particular, 6-month-old infants were presented with a set of images of cats and then tested on their looking behavior toward a set of pictures of dogs interspersed with novel cat pictures while ERPs were recorded. When tested, infants looked longer to the novel dogs than to the novel cats, suggesting they treated the within-category images (the novel cats) as more similar to the images from familiarization, despite the fact that all images were novel.

The ERPs to the pictures of cats and dogs also suggested that infants were sensitive to category information. In particular, the ERPs to the cats viewed early on during familiarization, as well as to the set of novel dog images, showed a negative slow wave (or a negative deflection of the ERP wave) between 1 and 1.5 s after the picture appeared. The decrease of these components to cats during the latter part of familiarization indicates that infants were beginning to respond to novel cats as though they were familiar. In other words, infants were responding to category-level information, not just item-level features. In addition, the dog pictures, but not the novel cat pictures, elicited a negative central component between 300 and 750, a component thought to reflect attentional allocation to novel stimuli in infants (Nelson, 1994). Quinn et al. suggest that this component may be related to the behavioral differences in looking to the dog and cat pictures, specifically the novelty preference for pictures of dogs observed after infants have been viewing only pictures of cats. Noting that the negative slow wave indexing categorization occurs after the component reflecting novelty detection, they suggest that recognizing similarity between within-category objects is a more complex process than recognition of novelty.

Using a similar procedure, Grossmann et al. (2009) also investigated the ERP signatures involved in infant categorization. The ERP recordings revealed a negative component at 300–500 ms in anterior cortical regions when infants were shown an image from a novel category, similar to the findings of Quinn et al. (2006). Grossmann et al. (2009) also found that novel exemplars of the familiarized category evoked a late positive component relative to the response to repeated items from the category. The authors suggest that this component may reflect updating of the category representation.

In sum, infants can track correlated properties of elements by the time they are 4 months of age. This ability plays a role in object perception, both in object segmentation and categorization, as well as in forming grammatical categories in the auditory domain. The data from the ERP recordings add to our understanding of the behavioral findings by showing real-time learning, helping to isolate different aspects of the process (responding to familiarity, novelty, and memory updating), and by shedding light on the neural processes underlying discrimination observed at test. In future studies, it will be important to test the neural correlates of learning novel categories, as current evidence pertains to categories that infants are familiar with prior to the experiment. In addition, it will be interesting to test whether there are parallels between the neural correlates observed in the visual domain and in grammatical category formation, as well as whether the signatures of category learning differ from those of learning sequential structure.

13.4 LEARNING ASSOCIATIONS BETWEEN WORDS AND REFERENTS

A central problem facing infant language learners is that of forming associations between words and their referents. One possibility is that this process begins no differently than the process of learning correlated object properties, as described in the previous section. Indeed, one might think of a label as simply another feature that is shared by objects within a category (Sloutsky and Robinson, 2008). Others have suggested that words are not simply a feature like any other but that they have a privileged role because they are inherently symbolic (Waxman and Gelman, 2009). On this view, words have a different relation to objects than do other features of objects, such as their appearance, from the very beginning of word learning. In either case, it is clear that learning the associations between words and their referents is a highly complex process, requiring a powerful but selective associative learning mechanism. Even the occurrence of a word with a very concrete and observable meaning, such as 'rabbit,' will coincide with many objects and events in the environment, sometimes, but not always, including an actual rabbit (e.g., Quine, 1960). Thus, establishing the referent of a novel word poses a formidable challenge. Even if the infant is fixating on a rabbit, she still must determine that the label refers to the animal itself, as opposed to its floppy ears, or to its color. Thus, the challenge facing the infant is to form an association between the word and some aspect of the environment that contains the referent, but that is neither overly broad nor too narrow.

While the demands of learning words are substantial, recent research suggests that infants' ability to form associations is both remarkably powerful and highly selective. Infants' sensitivity to the reliability of co-occurrence relationships between words and particular aspects of the environment may facilitate learning in such complex environments. Smith and Yu (2008) tested whether infants in the early stages of word learning can capitalize on these relationships using a cross-situational learning paradigm. They presented 12- and 14-month-old infants with six words whose referents were embedded in complex scenes. On each trial, infants heard a label (e.g., 'bosa') while viewing a scene consisting of the referent along with a distractor object. Given just a single trial of this nature, infants would be unable to determine which object was the referent. However, on other trials in which 'bosa' was presented, one of the same objects was presented along with a different distractor. On other trials, the label 'manu' was presented, and the object that

consistently occurred with it served as a distractor on other trials. Thus across trials, each label consistently occurred with one object. After infants were familiarized with the label–object training trials, they were tested using a preferential looking procedure. On each trial, infants were shown two objects simultaneously while the label for one of them was repeated. Infants looked significantly longer toward the object that matched the label at both ages, with stronger effects for the 14-month-olds than for the 12-month-olds.

These findings suggest that infants are sensitive to the reliability of label–object pairings across occurrences and can use information gathered across trials to settle on the most probable referent. However, it is important to note that words often occur in the absence of their referent and vice versa. Voloumanos and Werker (2009) thus tested whether infants learn label–object associations under more stochastic conditions. They presented 18-month-old infants with three word–object pairings, with each object labeled 10 times. One of the objects was labeled by the same word all 10 times, but the other two objects were labeled stochastically: most of the time, they occurred with one label (8 times), but on two occurrences, they were paired with a label that occurred predominantly with the other object. Infants were then tested on how well they learned the associations using a preferential looking task. The results suggest that infants were able to find the referent of words labeled stochastically: performance for word labeled 10 times or 8 times was the same when the correct referent was paired with a distractor object that had never occurred with the label. However, when a word labeled 8 times was tested with the correct referent (the object it had co-occurred with 8 times) paired with a distractor object that it had co-occurred with 2 times, infants' performance was at chance. Thus, experiencing the distractor object paired with the label just twice over the course of training disrupted recognition of the more frequent word/object pairing. Together, these findings suggest that probabilistic co-occurrences between labels and objects are harder for infants to learn than deterministic pairings, but that infants nonetheless develop some sensitivity to the association.

Recent work has shed light on the neurophysiological correlates of detecting reliable word–object associations during infancy. Friedrich and Friederici (2008) recorded ERPs while 14-month-olds were presented with two novel nonsense words, each occurring 8 times. One word was always presented with the same novel object (consistent referent), while the other was presented with a new object on each occurrence (inconsistent referent). After the first four exposures to each word, they found an early fronto-laterally localized negative component (N200-500) that was greater for the word with a consistent referent than for the word that had inconsistent

referents. Over the course of the second set of four exposures, this component diminished, and the words with consistent referents began to evoke an N400 in parietal regions. Infants were also tested on their memory for the consistent pairings 24 h later. On these trials, those words were paired with either the correct or incorrect referent. A larger N400 was observed when the words were presented with an incorrect referent.

The authors suggest that the N200-500 reflects early-occurring associative processes that begin to link together the auditory and visual features of the consistent events, and that the N400 reflects semantic integration: a stronger encoding of and memory for the word-referent relation. This interpretation is consistent with the finding that when 14-month-olds are presented with familiar words from their own language, there is a larger N400 to words presented with incongruous labels than congruous ones (Friedrich and Friederici, 2005). Taken together, these findings suggest that infants are sensitive to consistent co-occurrence relationships between words and objects, and that distinctive neural processes may underlie the formation of these referential associations.

13.4.1 Sensitivity to the Statistics of Objects Taking the Same Label

Another way infants' associations become more refined is in terms of the kinds of referents they associate with novel words. While infants are initially flexible in the kinds of referents they consider, recent studies suggest that this process rapidly becomes constrained by prior learning. Smith and colleagues (Colunga and Smith, 2003, 2005; Jones and Smith, 2002; Smith et al., 2002) suggest while learning novel words may initially require repeated associations between objects and labels, once a critical mass of labels has been acquired, they detect statistical regularities that characterize those associations. For example, object labels such as 'cup,' 'ball,' and 'phone' refer to a set of objects that have a common shape, and many of the words in children's early vocabularies are objects that are organized by shape (Samuelson and Smith, 1999). However, as children learn more words, a higher-order abstraction emerges from these specific associations: object labels tend to pick out groups of things with similar shapes. Such a generalization would allow children to extend object labels on the basis of shape as opposed to other features, such as color or texture. As a result, upon hearing the label 'crayon' referring to a red crayon, they can use it to refer to new crayon-shaped objects regardless of their color.

Based on these considerations, Smith et al. (2002) hypothesized that teaching children words from shape-based categories should result in increased word learning. They tested this hypothesis by bringing

17-month-old infants, who do not yet show a systematic shape bias, into the lab several times over 2 months, and teaching them names for novel objects. In one condition, the object categories were shape based: objects given the same names had similar shapes. In a control condition, objects that shared names did not have shape in common, but rather had some other property, such as their material, in common. Infants taught shape-based categories extended trained words to novel items based on shape, while infants in the nonshape-based conditions, or given no training, did not. Moreover, infants in the shape-based category training rapidly extended novel labels based on shape as well and showed a greater increase in their overall vocabularies over the 2-month training period than infants in the other conditions. These findings suggest that infants' experience with well-structured label–object pairings facilitates the rapid formation of new label–object associations, and the appropriate extension of those labels. Critically, this process results in rapid gains in vocabulary size and word-learning skill.

Just as increases in vocabulary size lead to more efficient word learning (Smith et al., 2002), the neurophysiological processes underlying word learning also appear to change as infants' vocabulary size increases. Torkildsen et al. (2008b) investigated the ERP signatures of 20-month-old infants in a novel word-learning task as a function of whether infants had fewer or more than 75 words in their productive vocabularies, a period that typically precedes a period of rapid vocabulary growth. In this task, infants were exposed to training blocks in which they viewed three familiar pictures, each labeled by the correct referent. They also saw three pictures of imaginary animals, each paired with a nonsense word. After each word had been presented with its referent five times, the pictures were presented with incongruous labels. All infants showed a larger N400 to the pairings of real words with incongruous objects at central and parietal recording sites, suggesting that the infants recognized the mismatch between label and referent. For the newly trained words, only the high-vocabulary infants showed an N400 when they were paired with an incongruent referent. This suggests that the low-vocabulary infants either did not learn the words or were able to form only a very weak association. The high-vocabulary infants showed more robust evidence of learning the novel mappings, though the N400 response was more broadly distributed than for the familiar words, suggesting that their sensitivity to the novel words differed from that to more established word–referent pairings.

Torkildsen et al. (2008a) also examined the ERPs during the five training trials. Consistent with the fact that only high-vocabulary infants showed an N400 indicative of successful learning at test, they found that only the high-vocabulary infants showed an increase in an early

negative component (N200-400) as the novel words repeated during training. This component has previously been associated with the early stages of learning the meaning of a word (see Friedrich and Friederici (2008)). While these findings do not suggest a neural mechanism by which infants narrow in on specific features as they gain skill in word learning, they do confirm that the process becomes more efficient as infants' vocabulary grows.

13.4.2 Sensitivity to the Internal Statistics of Labels in Word Learning

Additional constraints on the associations that infants are likely to form during word learning come from the properties of potential labels themselves. Graf Estes et al. (2007) tested the hypothesis that experience with statistical cues to word boundaries influences infants' ability to learn word–object pairings. In this study, 16-month-old infants were first familiarized with a stream of syllables in which the only cues to word boundaries were the TPs between adjacent syllables. Infants were then habituated to two label–object pairs. For some infants, the labels corresponded to the words (TP = 1.0), and for the other infants, the labels were equally frequent part-words (TP = 0.33) from the speech stream. After habituation to the label–object pairings, infants were tested using a Same-Switch procedure (Werker et al., 1998). The Same trials preserved the label–object pairings present during habituation, while the label–object pairings were switched on Switch trials. Learning is measured by the degree of increased looking to Switch trials, which is evidence of dishabituation. Only infants for whom the labels corresponded to words showed longer looking to the Switch trials, suggesting that learning words composed of syllables with high TPs is easier than learning labels that had internal statistics suggesting that the syllables did not cohere into a word. Likewise, Graf Estes et al. (2010) found that infants learn word–object associations better when the labels conform to the predominant phonotactic patterns of the native language. Experience with regularities in fluent speech may thus facilitate word learning by providing infants with candidate sound sequences to map to referents, and by promoting more accurate processing of those sound sequences, facilitating subsequent recall.

Data from ERP recordings in word-learning tasks also suggest that semantic processing is influenced by phonotactic information. As previously discussed, between 12 and 14 months of age, infants show a larger N400 when familiar words are presented with incongruous referents. Friedrich and Friederici (2005) found that both 12- and 19-month-old infants showed an early negative component that differed between real and

phonotactically legal nonsense words and phonotactically illegal nonsense words presented without referents, which likely reflects sensitivity to the relative familiarity of the words. At 19 months, infants also show an N400 response when novel words that are phonotactically legal in their language are presented with objects that have known labels, suggesting they recognize the label–object mismatch. However, phonotactically illegal words did not elicit an N400, suggesting that they may not have been perceived as potential labels. These findings are consistent with the hypothesis that the N400 response reflects an advance or change in word-learning skill, and specifically that phonotactic information plays an important role by the end of the second year.

Another potential source of information guiding word learning is the overlap between words' statistical and semantic properties at the level of grammatical categories. Words from different lexical categories are correlated with different semantic regularities (e.g., nouns tend to refer to objects and people, adjectives to properties such as color or texture, and verbs to actions or events). As previously discussed, words from different lexical categories can also be distinguished by a constellation of statistical cues, including distributional and phonological regularities (e.g., Christiansen et al., 2009; Kelly, 1992; Mintz et al., 2002; Monaghan et al., 2005).

A recent study tested whether infants can capitalize on experience with such cues in a word-learning task (Lany and Saffran, 2010). In this study, 22-month-old infants first listened to an artificial language that contained two word categories, disyllabic *X-words* and monosyllabic *Y-words*. For infants in the experimental group, phrases took the form *aX* and *bY*, and thus words from the *X* and *Y* categories were reliably marked by correlated distributional and phonological cues. Infants in the control group also heard *aX* and *bY* phrases, but in addition, they heard an equal number of *aY* and *bX* phrases. Thus, for the control infants, the word categories were not marked by correlated cues. Infants were then trained on pairings between phrases from the language and pictures of unfamiliar animals and vehicles. For both experimental and control groups, familiar *aX* phrases were paired with animal pictures, and familiar *bY* phrases were paired with vehicle pictures. Infants were then tested using a preferential looking procedure.

Interestingly, only the experimental infants were able to learn the trained associations between phrases and pictures, despite the fact that control infants had the same amount of experience with them. Moreover, only the experimental infants successfully generalized to novel pairings: when hearing a word with distributional and phonological properties of other words referring to animals, they mapped the word to a novel animal over a novel vehicle. These findings suggest that infants' experience with statistical cues marking word categories lays

an important foundation for learning the meanings of those words.

In sum, infants' ability to form associations is remarkably powerful, but it is also selective. Infants do not simply form an association between a word and object that happen to co-occur, but rather they form an association only when that object reliably occurs across other presentations of that word. And, while infants are initially relatively flexible in the kinds of associations they form between words and referents, recent studies suggest that this process rapidly becomes constrained by prior learning. For example, by 17 months of age, words with good sequential statistics are more readily associated with referents than sound sequences that do not have the characteristics of typical words in that language (Graf Estes et al., 2007). Moreover, once infants have begun to learn associations between words and referents, this knowledge influences the associations that infants will subsequently form. For example, infants rapidly begin to associate novel nouns with objects' shape over objects' color (e.g., Smith et al., 2002). Likewise, words that have statistical properties of a particular category are readily mapped to new referents from that category (Lany and Saffran, 2010).

13.5 CONCLUSIONS

The research reviewed in this chapter suggests that infants' sensitivity to statistical structure is powerful and nuanced and plays a role in diverse aspects of learning over the first several months of life. We have discussed infants' sensitivity to four different kinds of statistical structure: probability distributions, sequential structure, correlations between stimulus dimensions, and associations between different forms of information. While we have discussed these types of structure separately, these are likely not independent learning processes. For example, there are important similarities between learning sequential structure in word segmentation and in learning correlational structure in object segmentation. Tracking sequential structure within words may share important properties with learning sequential regularities between words and word categories. Likewise, the processes involved in learning word–label associations may be similar to those in learning feature co-occurrences in object categorization. Furthermore, real-world learning processes tap a combination of these learning mechanisms – for example, learning word–object associations critically relies on a complex interaction of sensitivity to phonotactic regularities, word segmentation, object categorization, word-order patterns, and grammatical categories.

A related point is that statistical learning mechanisms are tuned through experience. For example, in the

auditory domain, infants' sensitivity to reliable TPs between syllables facilitates segmenting words from fluent speech. In turn, this makes it possible to learn phonological regularities correlated with TPs, which may further facilitate segmentation in natural speech (Sahni et al., 2010). Likewise, segmentation facilitates learning word-order patterns, and tracking correlations between such distributional properties of words and their phonological properties can facilitate learning grammatical structure. In the domain of visual perception, tracking the co-occurrence of features may facilitate segmenting objects within complex scenes (Fiser and Aslin, 2002), and tracking how these features co-occur across instances facilitates the formation of object categories (Younger, 1985). In other words, as learners track simple structure, they build a foundation for tracking more complex, higher-order structure. They also appear to highlight relevant structure, which can facilitate learning more computationally challenging regularities (Lany and Gómez, 2008).

These studies also raise many intriguing questions for future research. For example, many studies manipulate the statistical properties of infants' input and show that this influences learning, but this does not suggest that infants are tracking input statistics veridically. Indeed, studies with older children suggest that they, unlike adults, tend to distort input statistics, maximizing probabilities rather than matching them (e.g., Hudson Kam and Newport, 2005). It also leads to questions of how infants encode these statistical regularities and how fine-grained their sensitivity is. Also, learning inherently entails perceiving, remembering, acting on the environment, and interacting with others, and thus in future research, it will be important to relate infants' statistical learning abilities to other developmental processes, such as perception, memory, and social interaction.

Another important area for future research is the underlying neurophysiology of statistical learning. Researchers have just begun to investigate the neurophysiological substrates of the behavioral findings, namely, how ERPs to stimuli consistent with familiarized patterns differ from ERPs to stimuli that are inconsistent. Given the challenges inherent in ERP research in infants, much of this work has tested learning in adults. However, recent advances in neurophysiological methodology in infants suggest that these techniques hold much promise for investigating learning mechanisms directly. Indeed, while the discrimination tasks typically used in infant behavioral studies can only probe the endpoint (or other static snapshot) of the learning process, psychophysiological techniques are particularly well suited to observing the process of learning.

In sum, infants' environments are rich with meaningful statistical structure, and infants readily track it. Far from being limited to simple associative learning, such as the stimulus–response learning of classical learning theory, the studies reviewed in this chapter reveal a sensitivity to statistical structure that is powerful and nuanced, capable of not only tracking fine-grained detail but also forming generalizations. Advances in the study of statistical learning mechanisms have shed new light on many aspects of development, such as auditory and visual perception, language development, and event processing. Continued study of infants' statistical learning mechanisms holds great promise for advancing the study of early cognitive development.

Acknowledgments

References

Abla, D., Okanoya, K., 2009. Visual statistical learning of shape sequences: An ERP study. Neuroscience Research 64, 185–190.

Abla, D., Katahira, K., Okanoya, K., 2008. Online assessment of statistical learning by event-related potentials. Journal of Cognitive Neuroscience 20, 952–964.

Aslin, R.N., Pisoni, D.B., Hennessy, B.I., Perey, A.J., 1981. Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. Child Development 52, 1135–1145.

Aslin, R.N., Jusczyk, P.W., Pisoni, D.B., 1998a. Speech and auditory processing during infancy: Constraints on and precursors to language. In: Kuhn, D., Siegler, R. (Eds.), Handbook of Child Psychology: Cognition, Perception, and Language, vol. 2. Wiley, New York, pp. 147–254.

Aslin, R.N., Saffran, J.R., Newport, E.L., 1998b. Computation of conditional probability statistics by 8-month-old infants. Psychological Science 9, 321–324.

Baldwin, D., Andersson, A., Saffran, J.R., Meyer, M., 2008. Segmenting dynamic human action via statistical structure. Cognition 106, 1382–1407.

Bomba, P.C., Siqueland, E.R., 1983. The nature and structure of infant form categories. Journal of Experimental Child Psychology 35, 609–636.

Braine, M.D.S., 1987. What is learned in acquiring words classes: A step toward acquisition theory. In: MacWhinney, B. (Ed.), Mechanisms of Language Acquisition. Erlbaum, Hillsdale, NJ, pp. 65–87.

Burnham, D., Kitamura, C., Vollmer-Conna, U., 2002. What's new pussycat? On talking to babies and animals. Science 296, 1435.

Cartwright, T.A., Brent, M.R., 1997. Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. Cognition 63, 121–170.

Chambers, K.E., Onishi, K.H., Fisher, C., 2003. Infants learn phonotactic regularities form brief auditory experience. Cognition 87, B69–B77.

Cheour, M., Ceponiene, R., Lehtokoski, A., et al., 1998. Development of language-specific phoneme representations in the infant brain. Nature Neuroscience 1, 351–353.

Christiansen, M.H., Onnis, L., Hockema, S.A., 2009. The secret is in the sound: From unsegmented speech to lexical categories. Developmental Science 12, 388–395.

Colunga, E., Smith, L.B., 2003. The emergence of abstract ideas: Evidence from networks and babies. Philosophical Transactions by the Royal Society 358, 1205–1214.

Colunga, E., Smith, L.B., 2005. From the Lexicon to expectations about kinds: A role for associative learning. Psychological Review 112, 347–382.

Cunillera, T., Toro, J.M., Sebastian-Galles, N., Rodriguez-Fornells, A., 2006. The effects of stress and statistical cues on continuous speech segmentation: An event-related brain potential study. Brain Research 1123, 168–178.

Cunillera, T., Càmara, E., Toro, J.M., et al., 2009. Time course and functional neuroanatomy of speech segmentation in adults. NeuroImage 48, 541–553.

Eimas, P.D., Quinn, P.C., 1994. Studies on the formation of perceptually based basic-level categories in young infants. Child Development 65, 903–917.

Eimas, P.D., Siqueland, E.R., Jusczyk, P., Vigorito, J., 1971. Speech perception in infants. Science 171, 303–306.

Fagan, J.F., 1970. Memory in the infant. Journal of Experimental Child Psychology 9, 217–226.

Fantz, R.L., 1964. Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. Science 146, 668–670.

Farmer, T.H., Christiansen, M.H., Monaghan, P., 2006. Phonological typicality influences on-line sentence comprehension. Proceedings of the National Academy of Sciences of the United States of America 103, 12203–12208.

Fiser, J., Aslin, R.N., 2002. Statistical learning of new visual feature combinations by infants. Proceedings of the National Academy of Sciences of the United States of America 99, 15822–15826.

Friederici, A.F., Wessels, J.M.I., 1993. Phonotactic knowledge of word boundaries and its use in infant speech perception. Perception & Psychophysics 54, 287–295.

Friedrich, M., Friederici, A.D., 2005. Phonotactic knowledge and lexical–semantic processing in one-year-olds: Brain responses to words and nonsense words in picture contexts. Journal of Cognitive Neuroscience 17, 1785–1802.

Friedrich, M., Friederici, A.D., 2008. Neurophysiological correlates of online word learning in 14-month-old infants. Cognitive Neuroscience and Neuropsychology 19, 1757–1761.

Gerken, L.A., Wilson, R., Lewis, W., 2005. Seventeen-month-olds can use distributional cues to form syntactic categories. Journal of Child Language 32, 249–268.

Gomez, R.L., 2002. Variability and detection of invariant structure. Psychological Science 13, 431–436.

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

Gómez, R.L., Lakusta, L., 2004. A first step in form-based category abstraction in 12-month-old infants. Developmental Science 7, 567–580.

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

Graf Estes, K., Evans, J.L., Alibali, M.W., Saffran, J.R., 2007. Can infants map meaning to newly segmented words? Psychological Science 18, 254–260.

Graf Estes, K., Edwards, J., Saffran, J.R., 2010. Phonotactic constraints on infant word learning. Infancy 16, 180–197.

Grossmann, T., Gliga, T., Johnson, M.H., Mareschal, D., 2009. The neural basis of perceptual category learning in human infants. Journal of Cognitive Neuroscience 21, 2276–2286.

Hudson Kam, C.L., Newport, E.L., 2005. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. Language Learning and Development 1, 151–195.

James, W., 1890. The Principles of Psychology. Holt, New York.

Johnson, E.K., Jusczyk, P.W., 2001. Word segmentation by 8-month-olds: When speech cues count more than statistics. Journal of Memory and Language 44, 548–567.

Johnson, E.K., Tyler, M.D., 2010. Testing the limits of statistical learning for word segmentation. Developmental Science 13, 339–345.

Jones, S.S., Smith, L.B., 2002. How children know the relevant properties for generalizing object names. Developmental Science 5, 219–232.

Jusczyk, P.W., Friederici, A.D., Wessels, J.M., Svenkerud, V.Y., Jusczyk, A.M., 1993. Infants' sensitivity to the sound patterns of native language words. Journal of Memory and Language 32, 402–420.

Jusczyk, P.W., Luce, P.A., Charles-Luce, J., 1994. Infants' sensitivity to phonotactic patterns in their native language. Journal of Memory and Language 33, 630–645.

Kellmann, P.J., Arterberry, M.E., 2000. The Cradle of Knowledge. The MIT Press, Cambridge, MA.

Kelly, M.H., 1992. Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. Psychological Review 99, 349–364.

Kirkham, N.Z., Slemmer, J.A., Johnson, S.P., 2002. Visual statistical learning in infancy: Evidence for a domain general learning mechanism. Cognition 83, B35–B42.

Kirkham, N.Z., Slemmer, J.A., Richardson, D.C., Johnson, S.P., 2007. Location, location, location: Development of spatiotemporal sequence learning in infancy. Child Development 78, 1559–1571.

Kluender, K.R., Lotto, A.J., Holt, L.L., Bloedel, S.L., 1998. Role of experience for language-specific functional mappings of vowel sounds. Journal of the Acoustical Society of America 104, 3568–3582.

Kuhl, P.K., 2004. Early language acquisition: Cracking the speech code. Nature Reviews Neuroscience 5, 831–843.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., Lindblom, B., 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. Science 255, 606–608.

Kuhl, P.K., Andruski, J.E., Chistovich, I.A., et al., 1997. Cross-language analysis of phonetic units in language addressed to infants. Science 277, 684–686.

Kuhl, P.K., Tsao, F.M., Liu, H.M., 2003. Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. Proceedings of the National Academy of Sciences of the United States of America 100, 9096–9101.

Kuhl, P.K., Stevents, E., Hayashi, A., Deguchi, T., Kiritani, S., Iverson, P., 2006. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. Developmental Science 9, F13–F21.

Lany, J., Gómez, R.L., 2008. Twelve-month-old infants benefit from prior experience in statistical learning. Psychological Science 19, 1247–1252.

Lany, J., Saffran, J.R., 2010. From statistics to meaning: Infant acquisition of lexical categories. Psychological Science 21, 284–291.

Lisker, L., Abramson, A.S., 1964. A cross-language study of voicing in initial stops: Acoustical measurements. Word 20, 384–482.

Liu, H.M., Kuhl, P.K., Tsao, F.M., 2003. An association between mothers' speech clarity and infants' speech discrimination skills. Developmental Science 6, F1–F10.

Marcovitch, S., Lewkowicz, D.L., 2009. Sequence learning in infancy: The independent contributions of conditional probability and pair frequency information. Developmental Science 12, 1020–1025.

Mareschal, D., Powell, D., Westermann, G., Volein, A., 2005. Evidence of rapid correlation-based perceptual category learning by 4-month-olds. Infant and Child Development 14, 445–457.

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

Maye, J., Weiss, D.J., Aslin, R.N., 2008. Statistical phonetic learning in infants: Facilitation and feature generalization. Developmental Science 11, 122–134.

McMurray, B., Aslin, R.N., Toscano, J.C., 2009. Statistical learning of phonetic categories: Insights from a computational approach. Developmental Science 12, 369–378.

Mintz, T.H., 2003. Frequent frames as a cue for grammatical categories in child directed speech. Cognition 90, 91–117.

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

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

Nelson, C.A., 1994. Neural correlates of recognition memory in the first postnatal year of life. In: Dawson, G., Fischer, K. (Eds.), Human Behavior and the Developing Brain. Guilford Press, New York, pp. 269–313.

Olson, G.M., Sherman, T., 1983. Attention, learning and memory in infants. In: Haith, M.M., Campos, J.J. (Eds.), Handbook of Child Psychology. Infancy and Developmental Psychobiology, vol. 2. Wiley, New York, pp. 1001–1080.

Pelucchi, B., Hay, J.F., Saffran, J.R., 2009a. Learning in reverse: Eight-month-old infants track backwards transitional probabilities. Cognition 11, 244–247.

Pelucchi, B., Hay, J.F., Saffran, J.R., 2009b. Statistical learning in a natural language by 8-month-old infants. Child Development 80, 674–685.

Quine, W.V.O., 1960. Word and Object. MIT Press, Cambridge, MA.

Quinn, P.C., Westerlund, A., Nelson, C.A., 2006. Neural markers of categorization in 6-month-old infants. Psychological Science 17, 59–66.

Redington, M., Chater, N., Finch, S., 1998. Distributional information: A powerful cue for acquiring syntactic categories. Cognitive Science 22, 425–469.

Rivera-Gaxiola, M., Klarman, L., Garcia-Sierra, A., Kuhl, P.K., 2005a. Neural patterns to speech and vocabulary growth in American infants. NeuroReport 16, 495–498.

Rivera-Gaxiola, M., Silva-Pereyra, J., Kuhl, P.K., 2005b. Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. Developmental Science 8, 162–172.

Rovee-Collier, C., 1986. The rise and fall of infant classical conditioning research: Its promise for the study of early development. Advances in Infancy Research 4, 139–159.

Saffran, J.R., Thiessen, E.D., 2003. Pattern induction by infant language learners. Developmental Psychology 39, 484–494.

Saffran, J.R., Wilson, D.P., 2003. From syllables to syntax: Multi-level statistical learning by 12-month-old infants. Infancy 4, 273–284.

Saffran, J.R., Aslin, R.N., Newport, E.L., 1996. Statistical learning by 8-month-old infants. Science 274, 1926–1928.

Saffran, J.R., Johnson, E.K., Aslin, R.N., Newport, E.L., 1999. Statistical learning of tone sequences by human infants and adults. Cognition 70, 27–52.

Sahni, S.D., Seidenberg, M.S., Saffran, J.R., 2010. Connecting cues: Overlapping regularities support cue discovery in infancy. Child Development 81, 727–736.

Sameroff, A.J., 1971. Can conditioned responses be established in the newborn infant: 1971? Developmental Psychology 5, 1–12.

Samuelson, L.K., Smith, L.B., 1999. Early noun vocabularies: Do ontology, category structure and syntax correspond? Cognition 73, 1–33.

Sanders, L.D., Newport, E.L., Neville, H.J., 2002. Segmenting nonsense: An event-related potential index of perceived onsets in continuous speech. Nature Neuroscience 5, 700–703.

Santelmann, L.M., Jusczyk, P.W., 1998. Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. Cognition 69, 105–134.

Sloutsky, V.M., Robinson, C.W., 2008. The role of words and sounds in infants' visual processing: From overshadowing to attentional tuning. Cognitive Science 32, 354–377.

Smith, K.H., 1969. Learning co-occurrence restrictions: Rule learning or rote learning? Journal of Verbal Behavior 8, 319–321.

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

Smith, L.B., Jones, S.S., Landau, B., Gershkoff-Stowe, L., Samuelson, L., 2002. Object name learning provides on-the-job training for attention. Psychological Science 13, 13–19.

Swingley, D., 2005. Statistical clustering and the contents of the infant vocabulary. Cognitive Psychology 50, 86–132.

Teglas, E., Girotto, V., Gonzales, M., Bonatti, L.L., 2007. Intuitions of probabilities shape expectations of the future at 12 months and beyond. Proceedings of the National Academy of Sciences of the United States of America 104, 19156–19159.

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

Theissen, E.D., Saffran, J.R., 2003. When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. Developmental Psychology 39, 706–716.

Theissen, E.D., Saffran, J.R., 2007. Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. Language Learning and Development 3, 73–100.

Torkildsen, J.V.K., Hansen, H.F., Svangstu, J.M., 2008a. Brain dynamics of word familiarization in 20-month-olds: Effects of productive vocabulary size. Brain and Language 108, 73–88.

Torkildsen, J.V.K., Svangstu, J.M., Hansen, H.F., 2008b. Productive vocabulary size predicts event-related potential correlates of fast-mapping in 20-month-olds. Journal of Cognitive Neuroscience 20, 1266–1282.

Valenza, E., Simion, F., Macchi Cassia, V., Umilta, C., 1996. Face preference at birth. Journal of Experimental Psychology. Human Perception and Performance 22, 892–903.

Valhaba, G.K., McClelland, J.L., Pons, F., Werker, J.F., Amano, S., 2007. Unsupervised learning of vowel categories from infant-directed speech. Proceedings of the National Academy of Sciences of the United States of America 104, 13273–13278.

Voloumanos, A., Werker, J.F., 2009. Infants' learning of novel words in a stochastic environment. Developmental Psychology 45, 1611–1617.

Vouloumanos, A., Hauser, M., Werker, J.F., Martin, A., 2010. The tuning of human neonates' preference for speech. Child Development 81, 517–527.

Waxman, S.R., Gelman, S.A., 2009. Early word-learning entails reference not merely associations. Trends in Cognitive Sciences 13, 258–263.

Werker, J.F., Tees, R.C., 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behaviour and Development 7, 49–63.

Werker, J.F., Cohen, L.B., Lloyd, V.L., Casasola, M., Stager, C.L., 1998. Acquisition of word-object associations by 14-month-old infants. Developmental Psychology 34, 1289–1309.

Werker, J.F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., Amano, S., 2007. Infant-directed speech supports phonetic category learning in English and Japanese. Cognition 103, 147–162.

Xu, F., Denison, S., 2009. Statistical inference and sensitivity to sampling in 11-month-old infants. Cognition 112, 97–104.

Xu, F., Garcia, V., 2008. Intuitive statistics by 8-month-old infants. Proceedings of the National Academy of Sciences of the United States of America 105, 5012–5015.

Yoshida, K.A., Pons, F., Maye, J., Werker, J.F., 2010. Distributional phonetic learning at 10 months of age. Infancy.

Younger, B.A., 1985. The segregation of items into categories by ten-month-old infants. Child Development 56, 1574–1583.

Younger, B.A., Cohen, L.B., 1986. Developmental changes in infants' perception of correlation among attributes. Child Development 57, 803–815.