Sensitivity to statistical information begets learning in early language development

Jessica F. Hay and Jill Lany

Growing interest in potential sources of statistical regularities in the environment, and in infants' ability to learn these regularities, has lead to a considerable amount of research examining the role of statistical learning in early development. One prevalent form of statistical information is co-occurrence. When two elements reliably co-occur, there is typically a meaningful relationship between those elements, and thus co-occurrence patterns potentially provide a very powerful source of information across different domains and modalities. Kirkham, Slemmer, and Johnson (2002) found that by 2 months of age, infants can detect sequential relationships in a series of visually-presented shapes, and Baldwin and colleagues (Baldwin, Andersson, Saffran, & Meyer, 2008) found that adults can use statistical information to learn regularities in dynamic action sequences. Also in the visual domain, 9-month-old infants can track elements that reliably co-occur across time and space, which tend to belong to the same object, and thus provide information about object boundaries (Fiser & Aslin, 2002).

Statistical information is also a potential cue to important aspects of language structure, and the extent to which sensitivity to co-occurrence regularities plays a role in infant language development is the focus of this chapter. One kind of co-occurrence relationship that may be particularly important, and that has been the focus of much research, is sequential relationships between syllables: 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). This type of sequential co-occurrence relationship is referred to as *conditional* or *transitional probability* (TP). Generally, transitional probability is the probability of one event given the occurrence of another event. This statistic refers to more than the frequency with which one element follows another, as it adjusts for the base rate of the first event or element. The TP of Y given X is represented by the following equation Eq. (1):

$$TP = P(Y|X) = \frac{frequency(XY)}{frequency(X)} \tag{1}$$

Take, for example, the following series of likely two word phrases that an infant might hear: *pretty baby, happy baby, cute baby, pretty kitty, pretty doggy, pretty horsey*. In this micro-corpus, every time an infant hears *ba* it is always followed by *by*. Therefore, within the word 'baby' the syllable sequence has a high transitional probability (TP = 1.0) and infants may therefore treat 'baby' as a good word candidate. Although the syllable *ba* predicts the syllable *by*, hearing *ty* (from 'pretty') does not necessarily predict *ba* (from 'baby'). In this micro-corpus, the transitional probability between syllable sequences *ty* and *ba* is 0.25, and therefore '*tyba*' (which crosses a word boundary) is a less good word candidate in English than 'baby'. Corpus analyses suggest that the TP between syllables are an imperfect but potentially useful cue to word boundaries in natural speech (Swingley, 2005; though, see Yang, 2004, for a different view), as transitional probabilities between syllables tend to be higher within words and lower across word boundaries. Thus, TP could serve as a word boundary cue for pre-lingual infants facing the challenge of segmenting words from fluent speech. Infants appear to be remarkably sensitive to statistical regularities beginning in early infancy. For example, Saffran, Aslin, & Newport (1996) found that 8-month-old infants can track these probabilities in an artificial language using synthesized speech, and Teinonen and colleagues (Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009) reported evidence of sensitivity to sequential statistics in speech from ERP recordings in neonates.

These and many other studies over the last 2 decades have yielded important insights about infants' ability to track statistical regularities in their environment (see Hay, 2009; Lany & Saffran, in press, for a review). However, many such studies have employed materials lacking the complexity of patterns found in the real world. For example, the artificial languages used in many experiments on word segmentation consist entirely of sequential regularities among a small number of syllables and lack the rich, multidimensional structure of natural language. Thus, it is unclear whether infants would still use statistical regularities when presented with natural language input that contains many potentially competing sources of information. Second, many of these studies present children with streams of speech that lack meaning. Because learning language entails learning semantic information in addition to phonological regularities, it is important to understand how statistical learning mechanisms contribute to other aspect of language development, such as word learning. Finally, another important goal is to understand how statistical learning at one level, such

as tracking co-occurrence relationships among syllables, might relate to learning about other aspects of language, such as the co-occurrence relationships between words and word categories that are important in grammatical structure.

In this chapter we will describe several recent studies that have begun to bridge the gap between studies suggesting that infants are highly adept at tracking statistical regularities in artificial languages with tasks that closer approximate the problems faced over the course of learning a natural language. Moreover, we will describe how sensitivity to relatively simple statistical regularities allows infants to learn new and increasingly complex dimensions of language structure, or how statistical learning begets learning.

1.   Statistical Learning in a Natural Language

The primary evidence supporting the existence of statistical learning mechanisms in infants comes from studies employing artificial language materials, typically miniature languages that have been designed to capture some aspect of structure found in natural language. For example, in the original statistical segmentation studies infants were presented with a synthetically produced speech stream, without pauses or other acoustic cues to word boundaries. The only available cue was a dip in the TPs (and other related sequential statistics, such as mutual information) between syllables and/or segments at word boundaries. Infants were then tested on their ability to discriminate sequences corresponding to *words* versus *nonwords* (syllables from the language assembled in a novel order), or a more subtle comparison, *words* versus *part-words* (syllable sequences spanning word boundaries, which can be matched for frequency in the speech stream; e.g., Aslin, Saffran, & Newport, 1998).

Although artificial language materials have been invaluable for the initial investigation of infant statistical learning mechanisms, it is obvious that such stimuli lack the complexity of a natural language on virtually every possible dimension. This problem of ecological validity has been acknowledged throughout the literature on infant statistical language learning. For example, the artificial languages tend to contain few words (typically just four to six), have a limited set of phonemes and syllables, lack other sequential regularities, rhythmic patterning, pitch changes, and other acoustic variability associated with natural languages. In addition,

words are repeated extremely frequently during exposure (45–90 times). Furthermore, there are typically no other sequential regularities present, such as those engendered by syntactic structure (though, see Saffran & Wilson, 2003).

For these reasons, the learning challenges presented by artificial languages are quite different than those presented by natural languages. In order to address this concern, researchers have systematically increased the complexity of artificial languages used in infant studies in several ways. For example, researchers have included naturally produced syllables instead of synthetically produced syllables (Graf Estes, Evans, Alibali, & Saffran, 2007; Sahni, Seidenberg, & Saffran, 2010), varied word length (Johnson & Tyler, 2010; Mersad & Nazzi, 2010; Thiessen, Hill, & Saffran, 2005) and stress cues (Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003; 2007) and provided multiple correlated cues to word boundaries (Sahni, et al., 2010). Generally, these studies have found that infants are still able to track statistical regularities in the input despite greater complexity (although, see Johnson & Tyler, 2010, for a counter-example), and that in some cases additional complexity facilitates statistical learning (e.g., Thiessen et al. 2005). However, while these miniature languages contain more structure than earlier artificial materials, they nonetheless lack the richness of natural languages.

In one early study examining infants' capacity to segment words from natural language, Jusczyk and Aslin (1995) presented 7.5-month-old infants with multiple repetitions of isolated words (such as *bike*, *cup*, *dog*, or *feet*) and then tested their listening preferences for sentences containing the word versus sentences without the word. In a second condition, infants were first familiarized with the sentences containing the target words and were then tested on their listening preferences for the isolated words. In both conditions, infants recognized the words in the novel context, suggesting that they must have noticed some similarity between the words in the sentences and the isolated words. Since transitional probability was not expressly manipulated, it is not clear whether mere familiarity with or statistical coherence in the target words was driving these effects.

Recently, however, Pelucchi and colleagues (Pelucchi, Hay, & Saffran, 2009a; 2009b), demonstrated that English-learning 8-month-old infants track statistical coherence, and specifically transitional probabilities, even when presented with complex language input that is naturally produced, grammatically correct, and semantically meaningful. Importantly, the natural language they used was not English, which English learning infants may

be able to segment using known words (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005), but instead was an unfamiliar natural language – Italian. The Italian corpora maintained virtually all of the complexities found in natural language, but the internal transitional probabilities between syllable sequences were expressly manipulated in a subset of the words.

To that end, Pelucchi et al.'s (2009a, Experiment 3) familiarization corpora consisted of grammatically correct and semantically meaningful standard Italian sentences, produced in a lively voice by a female native Italian speaker. The sentences contained 4 Italian target words with a strong/weak stress pattern: *fuga, melo, casa,* and *bici.* Importantly, two of the target words, *fuga* and *melo,* were *High Transitional Probability* (*HTP*) words as their component syllables never appeared without each other (i.e., *fu* never appeared in the absence of *ga,* and vice versa).

Thus, the TP of, for example, *fuga,* corresponds to:

$$TP(ga|fu) = \frac{f(fuga)}{f(fu)}$$

Because *fu* never appeared without *ga,* the internal transitional probability of *fuga* (and of *melo*) was 1.0. Two other words, *casa* and *bici,* were *Low Transitional Probability* (*LTP*) words as there were additional occurrences of the syllables *ca* and *bi,* all in strong (stressed) position. As a consequence, the TPs of *casa* and *bici* were .33. A counter-balanced language was also created to control for arbitrary listening preferences at test. Although the two pairs of words, *fuga/melo* and *casa/bici,* were equally frequent, they contained different internal TPs. In addition to sequential co-occurrence statistics of the target words, the Italian corpus contained much of the variability associated with natural language, including 23 different consonants, 7 vowels, over 100 unique syllables, a wide variety of syllable types (e.g., V, CV, VC, CCV), and varied word-length (i.e, 1-, 2-, 3-, and 4-syllable words), and also contained varied characteristic acoustic and prosodic patterns found in naturally spoken Italian. Despite the complexity of the natural language input, 8-month-old infants discriminated high vs. low transitional probability words following a very brief familiarization period (2 mins 15 sec).

In a second and related study, Pelucchi, Hay, and Saffran (2009b) investigated the directionality of the computation of sequential statistics. In principle, transitional probability can be defined in two different and symmetrical ways, depending on how the normalization step is performed. In addition to forward TPs (hereafter FTP) described in Eq. (1), it is also

possible to compute backward TPs (hereafter BTP), as shown in Eq. (2), measuring the likelihood of X **preceding** Y:

$$BTP = P(X \mid Y) = \frac{frequency(XY)}{frequency(Y)} \qquad (2)$$

Indeed, in a corpus analysis of English infant-directed speech, Swingley (1999) demonstrated that FTP and BTP are equally informative as independent cues to word boundaries.

In the Italian corpora used by Pelucchi et al. (2009a, Experiment 3), the transitional probabilities within the LTP words was lowered by including additional occurrences of the first syllable of each of those words. For example, the transitional probability of *casa* was lowered (to 0.33) by including other words that contained the syllable *ca*. However, the backward transitional probability remained high, at 1.0 because *sa* always preceded *ca*. Thus, Pelucchi et al. (2009b) presented another group of 8-month-old infants with an analogous familiarization corpus to the one described above, only this time the transitional probabilities within the LTP words was lowered by including additional occurrences of the second syllable of each of LTP words. Again, 8-month-old infants discriminated high vs. low transitional probability words following a very brief familiarization period. The Pelucchi et al. (2009a; 2009b) studies illuminate the power of statistical learning mechanisms and extend our understanding of their computational range; infants readily track both forward and backward TP even in complex, natural linguistic input.

Natural languages represent a noisy stimulus, in which the words of interest are interspersed amidst myriad other words, word repetitions are necessarily limited, and TPs are just one of many regularities in the input (e.g., prosodic patterns, morphological agreement, word order, etc.). These results thus provide a striking demonstration of statistical learning: infants detected sequential probabilities despite the substantial richness and complexity of the experimental materials. In fact, while such complexity has been considered to be a potential drawback of natural language materials, this complexity may prove to be advantageous for infant learners (for a discussion see Morgan, Shi, & Allopenna, 1996; Landau & Gleitman, 1985). Natural languages have the benefit of providing infants with multiple redundant cues to word boundaries, and are inherently more engaging than artificial languages. Indeed, prior research suggests that making artificial languages just a little more natural, by using infant-directed speech intona-

tion contours, can facilitate statistical learning (Thiessen et al, 2005). Similarly, infants learn more about word order when a word sequence is sung rather than spoken, providing engaging redundant cues to learning (Thiessen & Saffran, 2009).

2. The Role of Sequential Statistics in Linking Sounds to Meaning

Pelucchi et al.'s (2009a; 2009b) demonstration of statistical learning in a natural language allowed for greater ecological validity than previous experiments using artificial languages. However, these results – successful discrimination between test items that differed in transitional probability – tell us little about the representations that infants formed while listening to the fluent speech. For example, does the output of the segmentation task provide infants with something akin to words that are useful for subsequent language development? Infants may be able to use word segmentation processes to identify sound sequences that are likely to be individual words and thus ready to be associated with meanings. Alternatively, the representations that emerge from this statistical learning process may fail to intersect with subsequent word learning. On this latter view, statistical learning mechanisms, while available to infant language learners faced with fluent speech, do not render representations that can serve as potential words, thereby limiting the explanatory power of statistical learning accounts.

To begin to address these issues, Graf Estes et al. (2007) designed a method to study the connection between word segmentation and word learning. In this experiment, 17-month-old infants were first familiarized with an artificial language. The language consisted of four bisyllabic words, recorded so that the transitional probabilities between syllables were the only reliable cues to word boundaries. Infants were then presented with a word-object association task. They were taught two novel object-label associations where the labels were words from the artificial language (TP = 1.0), nonwords (sequences that did not appear in the language; TP = 0.0), or part-words (sequences that crossed word boundaries in the artificial language; TP = 0.5). Following habituation, infants were tested using a modified Switch paradigm (Werker, Cohen, Lloyd, Casasola, & Stager, 1998). Half of the test trials were Same trials, in which the object-label pairings presented during habituation remained unchanged. The other half of the test trials were Switch trials, in which the object and label pairings were switched such that Object A was paired with Label B and

vice versa. The results confirmed that the syllable co-occurrence information presented in the initial speech stream affected infants' subsequent word learning success: infants who were taught labels that were words in the speech stream showed significantly longer looking to the Switch trials than to the Same trials, indicating that they had learned the object-label pairings. However, infants who were taught labels that were nonwords or part-words from the language did not show differential looking to the two types of test trials. Only the high transitional probability sequences appeared to function as good candidate words. This pattern of results suggests that not only are infants able to track statistical regularities in sound sequences, but also that the output of this process can function as the input to future word learning.

Word learning tasks, combined with word segmentation tasks, provide a window into what infants learn through tracking statistical regularities in speech streams, because infants can use the output of statistical learning to support subsequent label-object association learning: one learning task feeds the next. However, like most statistical learning experiments, Graf Estes et al. (2007) used simple artificial language materials. Thus, it is not clear from the Graf Estes et al. (2007) study whether infants tracking regularities in natural languages are also able to form representations of sound sequences that are ready to link to meanings.

A more recent study by Hay and colleagues (Hay, Pelucchi, Graf Estes, & Saffran, 2011) examined whether infants can use the output of statistical learning in a natural language as the input to word learning. To that end, they combined the method developed by Graf Estes and colleagues (Graf Estes et al., 2007) with the Italian materials developed by Pelucchi et al. (2009a). In Experiment 1, 17-month-old infants were first familiarized with an Italian corpus of fluent speech, followed by a label-object association task. The crucial manipulation concerned the statistics of the labels relative to the Italian corpus. Half of the infants were trained on label-object pairings in which the labels were words from the corpus with a high internal transitional probability (HTP condition). The other half of the infants were trained on label-object pairings in which the labels had a low internal transitional probability (LTP condition). Importantly, both types of words occurred equally often in the Italian corpus. In both the HTP and LTP conditions, infants exhibited significantly longer looking times on Switch trials than on Same trials, suggesting that they readily learned both types of label-object pairings. The transitional probabilities internal to the labels did not appear to mediate learning; infants success-

fully mapped both HTP (TP = 1.0) and LTP (TP = 0.33) words as labels for objects.

By 17 months of age, infants are fairly skilled at learning new object-label associations, even in the absence of previous experience with the labels in fluent speech (e.g., Werker, Fennell, Corcoran, & Stager, 2002). It is thus possible that infants in Hay et al.'s (2011) first experiment did not need an extra boost from the word segmentation phase to successfully learn the labels. To rule out this possibility, Hay et al. (2011) ran a control experiment (Experiment 2) in which infants completed the label-object association task from Experiment 1 without any initial familiarization with the Italian corpus. They reasoned that if infants can learn the object-label associations based solely on their experience during the habituation phase of the label-object association task, then infants should also learn the label-object pairings in Experiment 2. In the absence of experience with the sequential statistics in the Italian corpus, infants were unsuccessful in learning to map the novel words to the novel objects. This suggests that infants in Experiment 1 needed the opportunity to segment the target words from fluent speech in order to map those target words onto novel objects.

An interesting puzzle emerges when comparing the label-learning patterns observed in Hay et al., (2011) with the findings of Graf Estes et al. (2007). In both studies, infants learned labels following segmentation experience. In Hay et al.'s (2011) Experiment 1, infants successfully mapped HTP and LTP Italian words as labels for objects. In contrast, Graf Estes et al. found that infants successfully learned high transitional probability sequences as object labels, but not low probability sequences that had occurred as part-words in an artificial language. One possible explanation for these diverging results is that the target words in Experiment 1 were trochees, whereas there were no prosodic cues present in the fluent speech used by Graf Estes et al. (2007). Thus, it is possible that the 17-month-old English-learning infants ignored the internal transitional probabilities of the target words and instead used a trochaic-based parsing strategy to segment words from the fluent speech.

Another possible explanation concerns the directionality of the computation of the sequential statistics. In the Italian corpora used by Hay et al. (2011), the transitional probabilities within the LTP words was lowered by including additional occurrences of the first syllable of each of those words. However, the backward transitional probability remained high, at 1.0. This manipulation suggests a potentially important difference from the low transitional probability part-word labels used by Graf Estes

et al. (2007), which contained low transitional probabilities in both directions (both forward and backward TPs = 0.5). It is thus possible that the high backward transitional probability of the LTP words played a crucial role in segmentation and subsequent word learning in Experiment 1.

In their third experiment, Hay et al. (2011) manipulated both forward and backward transitional probabilities in the LTP labels. As in Experiment 1, the HTP items contained high transitional probabilities (TPs = 1.0) in both directions. However, both the forward and backward transitional probabilities for the LTP words were lowered to 0.33 by adding further occurrences of the words' first <u>and</u> second syllables elsewhere in the corpus. Following familiarization with the Italian corpus, only infants habituated to HTP label-object pairings exhibited significantly longer looking times on Switch than Same trials; infants presented with LTP labels showed no difference in looking time. These results suggest that prior exposure to the fluent speech facilitated infants' ability to learn novel object labels, but only when the transitional probability between syllable sequences was high in at least one direction. When the forward and backward transitional probabilities were both lowered to 0.33, as in the LTP words used as labels in Hay et al.'s third experiment, infants failed to learn the label-object associations. Results from Hay et al.'s three experiments suggest that the internal cohesiveness of novel words, as measured by the strength of their internal transitional probabilities, influence how readily infants map these words to novel objects.

Recent research has also examined how real-world experience with phoneme co-occurrence affects subsequent word-learning. For example, Graf Estes, Edwards, & Saffran (2011) examined how experience with native-language phonotactic regularities affects subsequent word-learning. They exposed English-learning 18-month-old infants to 2 novel object-label parings consisting of either phonotactically legal sound sequences (i.e., sound sequences that were consistent with English phonotactics), *dref* and *sloob*, or 2 labels with phonotactically illegal sound sequences (i.e., sequences that are inconsistent with English phonotactics), *dlef* and *sroob*. Using a looking-while listening procedure (Fernald, Zangl, Portillo, & Marchman, 2008), Graf Estes and colleagues (Graf Estes et al., 2011) found that at test infants who were exposed to the phonotactically legal labels showed a larger increase in looking to the target objects than infants presented with phonotactically illegal labels. Consistent with these findings, Storkel (2001) has demonstrated that preschool-aged children (aged 3 to 6 years) are able to learn common sound sequence labels (i.e., labels with high phonotactic probability) with fewer exposures and retained

them with better accuracy than rare sequence labels (i.e., sequences with low phonotactic probability). These effects appear to be mediated by vocabulary size, suggesting that the more experience children have with native-language phonotactics, the greater effect phonotactics have on future word learning. Phonotactic probability also affects speech segmentation in infants as young as 9 months of age (e.g., Mattys & Jusczyk, 2001; Mattys, Jusczyk, Luce, & Morgan, 1999). Thus, experience with native-language phoneme co-occurrence regularities also appears to impact future language learning, in both younger and older children.

In sum, experience with sequential co-occurrence statistics present in natural language facilitates the identification of candidate words – words that are readied to be linked to meaning. Demonstrating that infants can track *and use* statistical regularities when faced with the complexity of natural language represents an important step in advancing our understanding of the role of statistics in natural language acquisition. Importantly, the output of speech segmentation feeds into learning increasing complex dimensions of language structure.

3. Statistical Learning and Grammatical Structure: Simple Grammatical Patterns

Thus far we have reviewed evidence suggesting that infants are sensitive to TPs in fluent synthesized speech, and even more impressively, in naturally-spoken Italian sentences. Moreover, this sensitivity begets subsequent word learning, such that sequences with high internal TPs are more readily mapped to visual referents. In addition to facilitating word segmentation and subsequent lexical development, sensitivity to co-occurrence information promotes learning other important dimensions of language structure, such as regularities in how words are combined into phrases and sentences. For example, Saffran & Wilson (2003) tested whether tracking TPs across syllables to find words would subsequently allow infants to track TPs across newly segmented words. They played 12-month-old infants synthesized speech strings in which TPs between syllables reliably cued word boundaries (i.e., within-words TPs between syllables were 1.0, and TPs of syllables spanning word boundaries were 0.25). However, the strings also contained regularities in how words were combined that could only be detected by tracking the TPs between the words. Infants were then tested on their ability to discriminate novel grammatical strings that were consistent with the word-order regularities in the training strings from ungrammatical

ones. Infants showed significant discrimination despite the fact that syllable-level TPs could not be used to distinguish between the grammatical and ungrammatical test strings. These findings suggest that sensitivity to co-occurrence relationships between syllables promotes learning sequential regularities at a higher level of language structure: segmenting words promotes learning statistical relationships between words.

Gomez and Gerken (1999) also found evidence that infants can use TPs to learn word order regularities within multi-word "sentences" in an artificial language. In their experiment, 12-month-old infants were familiarized to strings of nonsense words. As in natural language, word order was variable. In particular, the occurrence of a word did not predict the identity of the following word with 100% certainty, but some word sequences had higher TPs than others. Infants subsequently discriminated novel strings that contained high TPs from ungrammatical strings that contained lower TPs, indicating that they learned the probabilistic co-occurrence relationships between words.

These findings suggest that tracking TPs between words could play an important role in learning about higher-level language structure, and research with adults is consistent with this hypothesis. For example, Saffran (2001) found that adults can use these cues to learn phrase-structure information in an artificial grammar. Likewise, Thompson and Newport (2007) found that transitional probability cues within and across phrases are enhanced by common features of natural language such as phrasal movement and repetition, as well as the presence of optional phrases, and that such conditions promote adults' acquisition of phrase structure within an artificial language.

While the previous findings suggest that infants and adults can learn sequential relationships between words – indeed infants can do so by 12 months of age – learning the grammatical structure of a language requires learning how words from different grammatical categories can be combined. In other words, beyond tracking specific word sequences that are grammatical, such as *"the toy"*, infants and toddlers must learn more abstract patterns such as the fact that determiners like *"the"* and *"a"* precede nouns rather than follow them. This more abstract sensitivity is critical to grammatical development because it supports generalization to novel utterances.

How do infants begin to learn these aspects of their language? Several studies suggest that within a given language, the phonological properties of content words (e.g., nouns, verbs) and function words (e.g., determiners, pronouns, prepositions) are substantially different. For example, in both

English and Mandarin Chinese, content words are, on average, shorter, lower in amplitude, and contain simpler syllable structure, relative to function words (Morgan, Shi, & Allopenna, 1996). Amazingly, newborn infants can discriminate between content and function words, presumably on the basis of these cues (Shi, Werker, & Morgan, 1999). However, the specific phonological dimensions on which function and content words differ from language to language. In contrast, corpus studies of diverse languages (e.g., English, Korean, Italian) suggest that function words have a substantially higher token frequency than content words (Cutler & Carter, 1987; Gervain, Nespor, Mazuka, Horie, & Mehler, 2008; Kucera & Francis, 1967). These findings suggest that tracking word frequency might result in a rough distinction between these kinds of words.

Infants' sensitivity to this broad distinction between function and content words may promote subsequent language development by providing cues to simple grammatical patterns (Gervain et al., 2008). In particular, within a given language, phrases of different types tend to have similar underlying structure. For example, their corpus analysis of Japanese revealed that function words tend to occur in phrase-final positions, following content words, while function words tend to occur in phrase initial positions and are followed by content words in Italian. Gervain and colleagues found that by 8 months of age, infants prefer the patterning consistent with their native language, even when instantiated in an artificial language. These findings suggest that sensitivity to grammatical structure might be boot-strapped from experience with the ordering of function and content words at phrase boundaries.

Beyond forming coarse groupings of function words vs. content words, infants must form finer-grained grammatical categories, such as noun, verb, and determiner. One statistical cue that might promote forming these categories is referred to as a distributional cue, or regularities in the ordering of individual words that might point to commonalities among words. For example, nouns tend to occur after determiners, while verbs tend to follow pronouns and auxiliaries. Many studies have investigated the role of distributional cues in the acquisition of grammatical categories and their co-occurrence relationships using variants of an *aX bY* artificial language. These languages typically consist of the nonsense-word categories *a*, *b*, *X*, and *Y*, and restrictions on how words from these categories can be combined: *a*s precede *X*s but not *Y*s, and *b*s precede *Y*s but not *X*s, similar to determiner-noun and auxiliary-verb co-occurrence relationships in English. When the *X*s and *Y*s differ in their distributional properties alone, some learning is possible: learners encode and remember

the specific strings they were trained on, and also positional regularities, such as whether specific words occur in string-initial or string-final position. Critically, however, they fail to form word categories and detect their co-occurrence relationships (Smith, 1969). Interestingly, adults can use distributional information to form word categories when words from different categories were flanked by words providing distributional cues on either side, or "frequent frames", suggesting that distributional cues alone may be sufficient if they are particularly strong (Mintz, 2002).

Importantly, the presence of overlapping cues, such as correlations between words' distributional and phonological cues also facilitates learning the category-level co-occurrence relationships in adults (Frigo & McDonald, 1998; Braine, 1987). Natural languages incorporate multiple cues marking syntactic categories, such that words from different syntactic categories are distinguished by their distributional properties (Cartwright & Brent, 1997; Mintz, 2003; Mintz, Newport, & Bever, 2002; Monaghan, Chater, & Christiansen, 2005; Redington, Chater, & Finch, 1998), as well as by their phonological properties (Farmer, Christiansen, & Monaghan, 2006; Kelly, 1992; Monaghan et al., 2005). For example, nouns both tend to occur after "a" and "the", and have a strong-weak stress pattern. The fact that learners often benefit from the presence of these correlated cues in learning categories and their co-occurrence relationships is consistent with the possibility that the presence of correlated distributional and phonological cues facilitates learning by reducing computational and memory demands on learners. Rather than having to remember numerous individual aX or bY combinations to form word categories and learn their co-occurrence relationships, learners can track much simpler co-occurrence relationships between as and one phonological feature, and between bs and a different phonological feature. The relationships between as and bs and distinctive phonological features are referred to as marker-feature relationships because the as and bs resemble categories that serve as grammatical markers, as opposed to conveying semantic information.

To test whether infants can also use correlated cues to form grammatical categories, Gerken, Wilson, and Lewis (2005) exposed 17-month-old English-learning infants to Russian words from two categories that contained correlated cues marking their grammatical class. Specifically, Russian words often consist of a stem plus multiple grammatical morphemes marking case and grammatical gender. The familiarization set consisted of feminine words ending in the case markers "*oj*" and "*u*" and masculine words ending in the case markings "*ya*" and "*em*". The case markings provided distributional cues to the feminine or masculine cate-

gory membership of the stems. Critically, a phonological cue marking the category membership of the words was present on half of the words: three of the feminine words contained a derivational suffix "*k*", and 3 of the masculine words contained the suffix "*tel*". Thus, in many feminine words "*k*" was followed by "*oj*" and "*u*" (e.g., "*polkoj*" and "*polku*"), while in many masculine words "*tel*" was followed by "*ya*" and "*em*" (e.g., "*zhitelya*" and "*zhitelyem*"). Infants who were 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*". Using artificial language materials, Gomez & Lakusta, (2004) found that 12-month-olds can also learn correlations between distributional and phonological features when 100% and 87% but not 63% of *X*s and *Y*s contain distinctive phonological features. These findings suggest that tracking the correlations between distributional and phonological features of words plays an important role in learning grammatical patterns during infancy.

Statistical Learning and Grammatical Structure: Nonadjacent Dependencies

The studies of Gerken et al. (2005), Gomez & Gerken (1999), and Gomez & Lakusta (2004) suggest that infants can track predictive relationships between adjacent segments such as words and word categories. However, in natural languages, predictive relationships can also occur between nonadjacent elements, as in the relationship between auxiliaries such as "*is*" and the progressive inflection "*ing*" (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 may be more difficult take out than tracking adjacent dependencies because elements must be remembered long enough to be linked to other elements occurring later in time. Another challenge presented by nonadjacent dependencies is that there are many potentially irrelevant relationships for the learner to track, as dependent elements can be separated by several word elements.

Consistent with the high demands involved in learning nonadjacent dependencies, both infants and adults have substantial difficulty learning them, and succeed in doing so under highly restricted circumstances. For example, Newport and Aslin (2004) found that adults successfully detect

nonadjacecnt dependencies between similar segments (between consonants or vowels) but not between nonadjacent syllables. In contrast, (Pena, Bonatti, Nespor, & Mehler, 2002) found that adults used reliable transitional probabilities between nonadjacent syllables to segment words in a continuous speech stream. The participants did not, however, generalize, to novel words that maintained the nonadjacent dependencies but contained a novel middle syllable. These findings suggest that there may be some constraints on the kinds of nonadjacent dependencies that are readily acquired.

Tracking nonadjacent relationships appears to pose even more substantial challenges to infants. For example, while infants can track the relationships between adjacent elements in language-like materials well before they turn a year old (e.g., Saffran et al., 1996), they show evidence of sensitivity to grammatical relationships involving nonadjacent elements in their native language only at about 18 months of age (Santlemann & Jusczyk, 1998). Gomez (2002) found that 18-month-olds can also learn nonadjacent dependencies in artificial language materials, provided that potentially competing adjacent dependencies are highly variable and difficult to track. In subsequent research, Gomez & Maye (2005) found that 15-month-old infants also track nonadjacent dependencies, but that 12-month-olds fail to do so.

This developmental pattern suggests that increases in infants' memory capacity over the second year facilitates nonadjacent dependency learning. However, prior language experience may also play an important role in the development of this ability. Just as tracking TPs of adjacent syllables can also facilitate learning higher-order patterns in which those words are combined (Saffran & Wilson, 2003), Lany, Gómez, and Gerken (2007) found that exposing adults to adjacent dependencies greatly facilitates their ability to learn nonadjacent dependencies (see also Conway, Ellefson, & Christiansen, 2003; Elman, 1993; Newport, 1990; although see Rhode & Plaut, 1999). For infants to benefit from prior experience, they would have to generalize from strings containing adjacent dependencies to novel strings containing nonadjacent dependencies. Because infants can be more perceptually-bound than adults, there is active debate as to whether infants and children are capable of forming such generalizations (Tomasello, 2000; Gertner, Fisher, & Eisengart, 2006; Fisher, 2002). Thus, Lany & Gomez (2008) tested whether 12-month-olds infants given experience with adjacent dependencies could learn more difficult nonadjacent ones. They familiarized two groups of infants to strings from an *aX bY* artificial language as they playing quietly. Infants in the Experimental group heard *aX* and *bY* strings in which Xs and Ys differed in their phonological prop-

erties (*X*s were disyllabic and *Y*s were monosyllabic), and thus *X*- and *Y*-word categories were marked by correlated cues. Infants in the Control condition heard a mixture of *aX*, *aY*, *bX* and *bY* strings, such that *a*s and *b*s were followed by different sets of words (i.e., there were distributional cues present) but those words did not conform to any phonological regularities (i.e., both *a*s and *b*s predicted disyllabic and monosyllabic words). The Control language thus provided equivalent exposure to individual vocabulary elements and had the same intonational patterns and positional regularities as the *aX bY* language. Critically, however, it did not contain correlated cues marking word categories and their co-occurrence relationships. Infants in both conditions were subsequently habituated to strings containing nonadjacent *aX* and *bY* dependencies, but the particular dependencies had been withheld from familiarization. Only the Experimental infants subsequently dishabituated to strings containing violations of the nonadjacent dependencies. These results suggest that while 12-month-olds typically fail to track nonadjacent dependencies, given relevant prior experience with simpler adjacent dependencies, they can subsequently detect novel nonadjacent relationships between words from those categories.

Interestingly, Lany and Gomez (2008) found that only female infants were more likely to generalize to novel nonadjacent dependencies based on experience with adjacent ones. Willits, Lany, & Saffran (under review) also recently found evidence that at 24 months of age female infants are better able to generalize nonadjacency learning than male infants. This pattern of results is consistent with other findings of sex differences in language development. For example, females tend to have more words in their vocabularies than males of the same age (Nelson, 1973), and also tend to lead males in early word combinations (Schacter, Shore, Hodapp, Chalfin, & Bundy, 1978). Hartshorne and Ullman (2006) found that female toddlers are more likely to generalize the past-tense ending to irregular verbs (e.g., saying "holded" instead of "held"), particularly for irregular verbs sharing phonological features with regular verbs (e.g., the irregular verb "hold" is phonologically similar to the regular verbs "fold" and "mold"). They suggested that females formed stronger associations between phonologically-related verbs, and inappropriately generalized the regular past-tense morphology to irregular phonological neighbors as a result. These findings suggest that female infants may from stronger associations between phonologically-related words, and may shed light on the sex difference observed in Lany & Gomez (2008). If female infants are more sensitive to phonological similarities, they might be better able to notice similarities between *X*- and *Y*-elements based on their phonological

properties (i.e., the syllable-number cue), and subsequently to generalize to novel nonadjacent *aX* and *bY* combinations. Whether these findings reflect differences in phonological processing, memory development or other processes involved in language acquisition, they pose an intriguing question for further research on the underlying basis of sex differences in language development.

In sum, the findings from artificial language studies suggest that infants may begin to learn the distributional and phonological properties marking grammatical categories such as "noun" and "verb" and their co-occurrence relationships by the time they are 12-months old. Indeed, there is evidence that English-learning infants begin tracking distributional regularities by 12 months of age in their native language (Mintz, 2006), and by 14-months of age in German-learning infants (Hohle, Weissenborn, Kiefer, Schulz, & Schmitz, 2004). Thus, infants' experience with statistical cues appears to provide a critical foundation for learning adjacent relationships, as well as more difficult nonadjacent ones.

4. Linking Sounds to Meaning: Statistical Cues Marking Word Categories Promote Learning the Semantic Properties of Word Categories

Beyond promoting sensitivity to grammatical patterns, infants' experience with statistical regularities in the sound stream cueing grammatical patterns may provide a foundation for learning words' meanings. As previously mentioned, word-frequency may provide clues to grammatical structure, such as the broad distinction between function word and content words. Because content words such as "cat", "ball", "shoe", "run" and "crawl" tend to be mapped to referents, while function words such a "the", "on", or "that" do not have such concrete semantic referents, Hochmann, Endress, and Mehler (2010) hypothesized that tracking frequency information may also play a role in learning word meanings. To test this, they presented monolingual Italian-learning 17-month-olds with a corpus of unfamiliar French sentences exemplifying these frequency imbalances between function and content words. After listening to the familiarization corpus, infants were presented with a determiner-noun phrase while viewing a picture. The determiner had been presented much more frequently than the content word in the familiarization corpus. Infants were then tested on whether they formed a stronger association between the more frequent determiner or the less frequent content word. Despite the fact that the determiner and noun were paired with the picture equally often, infants appeared to form a selective association between the noun and

the picture. Infants who were exposed to the phrase-picture pairings with no exposure to the corpus containing information about the frequency imbalance showed no such preference. These findings suggest that when hearing multi-word phrases in the context of a potential referent, infants may be biased to map the relatively infrequent word to the referent.

Beyond sensitivity to which words are likely to be mapped to concrete semantic referents, statistical cues might facilitate forming a selective mapping between content words and referents by narrowing the pool of possible referents considered as likely candidates. In particular, words from different lexical categories are correlated with different semantic properties: e.g., nouns tend to refer to objects and people, adjectives to properties such as color or texture, and verbs to actions or events. Given the evidence that infants can use correlated statistical cues to form word categories, in a recent study Lany and Saffran (2010) tested whether they can also capitalize on experience with such cues to learn the semantic properties of categories. They first played 22-month-old infants an *aX bY* artificial language similar to the one used by Lany & Gomez (2008): Infants in the Experimental group heard *aX* and *bY* phrases, and thus for them the *X* and *Y* words differed in their distributional and phonological properties (i.e., membership in the *X* and *Y* categories was reliably marked by correlated phonological and distributional cues). In contrast, infants in the Control group were familiarized with *aX* and *bY* phrases in addition to *aY* and *bX* phrases, as in the study by Lany & Gomez (2008). Thus, in the Control group, words' phonological properties did not covary with their distributional properties, and the word categories were not marked by correlated cues. Infants in both conditions were then trained on pairings between phrases from the language and pictures of unfamiliar animals and vehicles. Critically, a subset of *aX* and *bY* phrases were heard by infants in both Experimental and Control groups, and these familiar *aX* phrases were paired with animal pictures, and familiar *bY* phrases were paired with vehicle pictures (or *vice versa*). Thus, for infants in both conditions there were reliable associations between individual *X*s and *Y*s and specific referents, and *aX* phrases always referred to referents of one kind or category, and *bY* phrases always referred to instances of the other category. Interestingly, only the Experimental infants learned the trained picture-phrase associations, 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 new word with both distributional and phonological properties of other words referring to animals, they mapped the word to a novel animal over a novel

vehicle. The findings of this experiment suggest that infants' experience with reliable cues in the initial listening phase promoted learning the semantic properties of individual words, as well as the generalization that aX phrases refer to animals and bY phrases to vehicles.

These findings are consistent with research on the role of sensitivity to correlated features in learning object categories. In particular, object categories (e.g., cups, dogs, birds, and trees) are structured such that properties or features characteristic of a category tend to co-occur within individual instances of the category. For example, the presence of one feature of a tree (e.g., branches) is correlated with the presence of the others (e.g., leaves, bark, and trunk) but less strongly correlated with the presence of features associated with objects from other categories (e.g., ceramic handles, fur, and beaks). Mareschal, Powell, Westermann, and Volein (2005) and Younger (1985) found that when the features of a set of fantastical animals form 2 clusters (e.g., fuzzy tails co-occur with ears, while vs. feathered tails co-occur with antlers), infants group the animals into categories corresponding to correlations between values on these dimensions. However, when feature values are randomly distributed, infants do not group the animals into separate categories. These findings suggest that there may be a parallel in the mechanisms underlying forming perceptually based object categories and forming grammatical categories. An important question for future research is whether mature representation of grammatical categories can be characterized in terms of sensitivity to correlated cues, or whether linguistic categories are ultimately organized in a fundamentally different way.

While the findings of Lany and Saffran (2010) suggest that infants readily learn intercorrelations between the distributional, phonological, and semantic properties of word categories, it is possible that the respective associations they formed between semantic information and each of the statistical cues could differ in strength. Thus, Lany and Saffran (2011) probed the role of each of these cues in the word learning process. Specifically, after listening to the artificial language containing correlated cues and being trained on associations between phrases and pictures as in Lany and Saffran (2010, Experimental Condition), infants were tested on their ability to find a novel referent given either only a distributional cue or only a phonological cue to category membership. Interestingly, infants' use of phonological and distributional cues to generalize was related to their native-language proficiency. Specifically, infants with smaller vocabularies successfully used phonological cues to generalize to the novel referents, but failed to use distributional cues. In contrast, infants with larger native-

language vocabularies and more advanced grammatical skills used distributional cues to words' category membership to generalize. These findings suggest that distributional and phonological information contribute independently to word-learning, and may also develop on different timetables. Given the evidence that children who are delayed in their vocabulary development, or "late talkers", as well as children with Specific Language Impairment, show deficits in using grammatical knowledge (frequent grammatical morphemes) to build their vocabularies (Moyle, Ellis Weismer, Evans, & Lindstrom, 2007), future work in this paradigm may shed light on the processes critical to their acquisition and successful use in word learning.

The findings of Lany and Saffran (2010; 2011) suggest that infants can use statistical information, and specifically distributional and phonological cues, to discover novel associations between word categories and semantic properties. According to the syntactic bootstrapping hypothesis (e.g., Landau & Gleitman, 1985), learning the meanings of words crucially relies on sensitivity to syntactic information, such as the syntactic frame in which a novel word is heard. The findings of Lany and Saffran suggest that experience with the distributional and phonological regularities marking words' category membership may play an important role in the development of the ability to use structural information to learn words. In future studies it will be important to pursue this link between statistical learning and syntactic bootstrapping by testing whether infants capitalize on statistical cues marking word categories in their native language. For example, given the evidence that English and German-learning infants are sensitive to correlated cues marking word categories like "noun" and "verb" by 12–14 months, this sensitivity could support subsequent word learning in natural language just as they support word-learning in an artificial language. Thus, an important question is whether statistical regularities play a role in the infants' use of syntactic bootstrapping in their native language. Work by Sandhofer and Smith (2007) suggests that this might be the case; in speech to infants and children adjectives are frequently used in ambiguous frames, and in fact are often used in frames that nouns can also be used in. Children have difficulty learning adjectives when they are presented in such ambiguous frames, but using novel adjectives in more distinctive frames (frames that distinguish them from nouns) can facilitate learning adjectives. An interesting possibility that should be addressed in future research is that statistical cues may play a role in the difficulties children have in learning other word classes, such as verbs.

Conclusions

Statistical learning plays an important role in learning structure in the sound stream, such as phonological regularities important for word segmentation as well as word-order regularities relevant to acquiring grammatical structure. Critically, sensitivity to statistical structure in one area of language can bootstrap infants' learning of other dimensions of language structure. For example, using TPs between syllables to segment fluent speech subsequently facilitates learning word-order patterns. Likewise, tracking correlations between such distributional properties of words and their phonological properties can facilitate learning complex grammatical structure. Moreover, tracking statistical regularities at all of these levels facilitates learning semantic information. By 17 months of age, words with good sequential statistics are more readily associated with referents than sound sequences that do not have characteristics of typical words in that language (Graf Estes et al., 2007; in press; (Hay et al., 2011)). Once infants have begun to learn associations between words and referents, this knowledge influences the associations that infants will subsequently form: Words that have statistical properties of a particular category are readily mapped to new referents from that category (Lany & Saffran, 2010; 2011). All together, these studies suggest that infants are able to track statistical regularities relevant to many aspects of language despite the considerable complexity of such structures, and shed important light on the mechanisms supporting language development.

References

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

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

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K.
>   2005   Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science*, 16, 298–304.

Braine, M.D.S.
>   1987   What is learned in acquiring word classes – A step toward an acquisition theory. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Conway, C. M., Ellefson, M. R. & Christiansen, M. H.
  2003   When less is less and when less is more: Starting small with
                staged input. In *Proceedings of the 25th Annual Conference
                of the Cognitive Science Society*, 270–275. Mahwah, NJ:
                Lawrence Erlbaum.

Cutler, A. & Carter, D. M.
  1987   The predominance of strong initial syllables in the English
                vocabulary. *Computer Speech and Language*, 2, 133–142.

Elman, J. L.
  1993   Learning and development in neural networks: The importance
                of starting small. *Cognition*, 48, 71–99.

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

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A.
  2008   Looking while listening: Using eye movements to monitor spoken
                language comprehension by infants and young children. In I. A.
                Sekerina, E. M. Fernandez, & H. Clahsen (Eds.), Developmental
                psycholinguistics: On-line methods in children's language process-
                ing (pp. 97–135). Amsterdam, Netherlands: John Benjamins
                Publishing Company.

Fisher, C.
  2002   The role of abstract syntactic knowledge in language acquisition:
                A reply to Tomasello (2000). *Cognition*, 82, 259–278.

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

Frigo, L. & McDonald, J. L.
  1998   Properties of phonological markers that affect the acquisition of
                gender-like subclasses. *Journal of Memory & Language*, 39, 218–
                245.

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

Gertner, Y., Fisher, C., & Eisengart, J.
  2006   Learning words and rules: Abstract knowledge of word order in
                early sentence comprehension. *Psychological Science*, 17, 684–
                691.

Gervain, J., Nespor, M., Mazuka, R., Horie, R., & Mehler, J.
  2008   Bootstrapping word order in prelexical infants: A Japanese–
                Italian cross-linguistic study. *Cognitive Psychology*, 57, 56–74.

Gómez, R. L.
  2002   Variability and detection of invariant structure. *Psychological Science*, 13(5), 431–436.
Gómez, R. L. & Gerken, L.
  1999   Artificial grammar learning by 1-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(2), 183–206.
Graf Estes, K., Edwards, J., & Saffran, J. R.
  2011   Phonotactic constraints on infant word learning. *Infancy*, 16, 180–197.
Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R.
  2007   Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, 18, 254–260.
Hartshorne, J. K. & Ullman M. T.
  2006   Why girls say 'holded' more than boys. *Developmental Science*, 9, 21–32.
Hay, J. F.
  2009   Statistical learning of language. In: Squire L.R. (Ed). Encyclopedia of Neuroscience, vol. 9, pp. 387–392. Oxford: Academic Press.
Hay, J. F., Pelucchi, B., Graf Estes, K., & Saffran, J. R.
  2011   Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology*, 63, 93–106.
Hochmann, J., Endress, A. D., & Mehler, J.
  2010   Word frequency as a cue for identifying function words in infancy. *Cognition*, 115, 444–457.
Hohle, B., Weissenborn, E., Kiefer, D., Schulz, A., & Schmitz, M.
  2004   Functional elements in infants' speech processing: The role of determiners in the syntactic categorization of lexical elements. *Infancy*, 5(3), 341–353.
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(2), 339–345.
Jusczyk, P. W. & Aslin, R. N.
  1995   Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, 29, 1–23.

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.

Kucera, H. & Francis, N.
1967   *A computational analysis of present-day American English.* Providence, RI: Brown University Press.

Landau, B. & Gleitman, L. R.
1985   Language and experience: Evidence from the blind child. Cambridge, MA: Harvard University Press.

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

Lany, J., Gómez, R. L., & Gerken, L.
2007   The role of prior experience in language acquisition. *Cognitive Science*, 31, 481–507.

Lany, J. & Saffran, J. R.
2010   From statistics to meanings: Infants' acquisition of lexical categories. *Psychological Science*, 21, 284–291.

Lany, J. & Saffran, J. R.
2011   Statistical learning mechanisms in infancy. In P. Rakik & J. Rubenstein (Eds.), *Developmental Neuroscience – Basic and Clinical Mechanisms.* Elsevier.

Lany, J. & Saffran, J. R.
2011   Interactions between statistical and semantic information in infant language development. *Developmental Science*, 14, 1207–1219.

Mattys, S. L. & Jusczyk, P. W.
2001   Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, 78, 91–121.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L.
1999   Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, 38, 465–494.

Mersad K. & Nazzi T.
2010   When mom helps statistics. Poster presented at the International Conference on Infant Studies, Baltimore, MD.

Mintz, T. H.
2002   Category induction from distributional cues in an artificial language. *Memory & Cognition*, 30, 678–686.

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

Mintz, T. H.
2006   Finding the verbs: distributional cues to categories available to young learners. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action Meets Word: How Children Learn Verbs*, p. 31–63. New York: Oxford University Press.

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 contribution of phonological and distributional cues in grammatical categorization. *Cognition*, 96, 143–182.

Morgan, J. L., Shi, R. & Allopenna, P.
1996   Perceptual bases of rudimentary grammatical categories. In J. L. Morgan & K. Demuth (Eds), *Signal to syntax*. Hillsdale, NJ: Erlbaum.

Moyle, M. J., Weismer, S. E., Evans, J. L., and Lindstrom, M. J.
2007   Longitudinal relationships between lexical and grammatical development in typical and late-talking children. *Journal of Speech, Language, and Hearing Research*, 50, 508–528.

Nelson, K.
1973   Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development*, 38.

Newport, E. L.
1990   Maturational constraints on language learning. *Cognitive Science*, 14, 11–28.

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

Pelucchi, B., Hay, J. F., & Saffran, J. R.
2009a   Statistical learning in a natural language by 8-month-old infants. *Child Development*, 80(3), 674–685.

Pelucchi B., Hay J. F., & Saffran J. R.,
2009b   Learning in reverse: Eight-month old infants track backward transitional probabilities. *Cognition*, 113, 244–247.

Pena, M., Bonatti, L. L., Nespor, M., & Mehler, J.
2002   Signal-driven computations in speech processing. *Science*, 298, 604–607.

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

Rohde, D. L. T. & Plaut, D. C.
1999   Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition*, 72, 67–109.

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

Saffran, J. R.
2001   The use of predictive dependencies in language learning. *Journal of Memory and Language*, 44, 493–515.

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

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

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

Sandhofer, C. & Smith, L. B.
2007   Learning adjectives in the real world: How learning nouns impedes learning adjectives. *Language, Learning and Development*, 3(3), 233–267.

Schacter, F. F., Shore, E., Hodapp, R., Chalfin, S., & Bundy, C.
1978   Do girls talk earlier?: Mean length of utterance in toddlers. *Developmental Psychology*, 13, 388–392.

Shi, R., Werker, J., & Morgan, J.
1999   Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, 72, B11–B21.

Smith, K. H.
1969   Learning co-occurrence restrictions: rule induction or rote learning? *Journal of Verbal Learning and Verbal Behavior*, 8(2), 319–321.

Storkel, H. L.
2001   Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research*, 44, 1321–1337.

Swingley, D.
1999   Conditional probability and word discovery: A corpus analysis of speech to infants. In *Proceedings of the 21st Annual Meeting of the Cognitive Science Society* (pp. 724–729). Mahwah, NJ.: LEA.

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

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(1), 21.

Thiessen, 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.

Thiessen E. D., Hill E. A., & Saffran J. R.
   2005   Infant-directed speech facilitates word segmentation. *Infancy*, 7(1), 53–71.
Thiessen E. D. & Saffran, J. R.
   2007   Learning to learn: Infants acquisition of stress-based strategies for word segmentation. *Language and Development*, 3, 73–100.
Thiessen, E. D. & Saffran, J.R.
   2009   How the melody facilitates the message and vice versa in infant learning and memory. *Annals of the New York Academy of Sciences*, 1169, 225–233.
Thompson, S. P. & Newport, E. L.
   2007   Statistical learning of syntax: The role of transitional probability. *Language Learning and Development*, 3, 1–42.
Tomasello, M.
   2000   Do young children have adult syntactic competence? *Cognition*, 74, 209–253.
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., Fennell, C. T., Corcoran, K. M., & Stager, C. L.
   2002   Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, 3(1), 1–30.
Willits, J. A., Lany, J., & Saffran, J. R.
   under review   Infants can use meaning to learn about higher order linguistic structure.
Yang, C.
   2004   Universal grammar, statistics, or both. *Trends in Cognitive Sciences*, 8, 451–456.
Younger, B. A.
   1985   The segregation of items into categories. *Child Development*, 56, 1574–1583.