PART II

Words, Language, and Music

Chapter 5

Connectionist Explorations of Multiple-Cue Integration in Syntax Acquisition

Morten H. Christiansen, Rick Dale, and Florencia Reali

Among the many feats of learning that children showcase in their development, syntactic abilities appear long before many other skills, such as riding bikes, tying shoes, or playing a musical instrument. This is achieved with little or no direct instruction, making it both impressive and even puzzling, because mastering natural language syntax is one of the most difficult learning tasks that humans face. One reason for this difficulty is a "chicken-and-egg" problem involved in acquiring syntax. Syntactic knowledge can be characterized by constraints governing the relationship between grammatical categories of words (such as noun and verb) in a sentence. At the same time, the syntactic constraints presuppose the grammatical categories in terms of which they are defined; and the validity of grammatical categories depends on how they support those same syntactic constraints. A similar "bootstrapping" problem faces a student learning an academic subject such as physics: understanding momentum or force presupposes some understanding of the physical laws in which they figure; yet these laws presuppose these very concepts. The bootstrapping problem solved by very young children seems much more daunting, both because the constraints governing natural language are so intricate, and because these children do not have the intellectual capacity or explicit instruction present in conventional academic settings. Determining how children accomplish the astonishing feat of language acquisition remains a key question in cognitive science.

By 12 months, infants are attuned to the phonological and prosodic regularities of their native language (Jusczyk, 1997; Kuhl, 1999). This perceptual attunement may provide an essential scaffolding for later learning by biasing children toward aspects of language input that are particularly informative for acquiring grammatical knowledge. In this chapter, we hypothesize that integrating multiple probabilistic cues (phonological, prosodic, and distributional) by perceptually attuned general-purpose learning mechanisms may hold promise for explaining how children solve the bootstrapping problem. Multiple cues can provide reliable evidence about linguistic structure that is unavailable from any single source of information.

In the remainder of this chapter, we first review empirical evidence suggesting that infants may use a combination of phonological, prosodic, and distributional cues to bootstrap into syntax. We then report a series of simulations demonstrating the computational efficacy of multiple-cue integration within a connectionist framework (for modeling of other aspects of cognitive development, see the chapter by Mareschal & Westermann, this volume). Simulation 1 shows how multiple-cue integration results in better, faster, and more uniform learning. Simulation 2 uses this initial model to mimic the effect of grammatical and prosodic manipulations in a sentence comprehension study with 2-year-olds (Shady & Gerken, 1999). Simulation 3 uses an idealized representation of prenatal exposure to gross-level phonological and prosodic cues, leading to facilitation of postnatal learning of syntax by the model. Simulation 4 demonstrates that adding additional distracting cues, irrelevant to the syntactic acquisition task, does not hinder learning. Finally, Simulation 5 scales up these initial simulations, showing that connectionist models can acquire aspects of syntactic structure from cues present in actual child-directed speech.

THE NEED FOR MULTIPLE LANGUAGE-INTERNAL CUES

In this section, we identify three kinds of constraints that may serve to help the language learner solve the syntactic bootstrapping problem. First, innate constraints in the form of linguistic universals may be available to discover to which grammatical category a word belongs, and how they function in syntactic rules. Second, language-external information, concerning observed semantic relationships between language and the world, could help map individual words onto their grammatical function. Finally, language-internal information, such as aspects of phonological, prosodic, and distributional patterns, may indicate the relation of various parts of language to each other, thus bootstrapping the child into the realm of syntactic relations. We discuss each of these potential constraints below, and conclude that some form of language-internal information is needed to break the circularity.

Although innate constraints likely play a role in language acquisition, they cannot solve the bootstrapping problem. Even with genetically prescribed abstract knowledge of grammatical categories and syntactic rules (e.g., Pinker, 1984), the problem remains: Innate knowledge requires building in universal mappings across languages, but the relationships between words and grammatical categories clearly differ cross-linguistically (e.g., the sound /su/ is a noun in French (*sou*) but a verb in English (*sue*)). Even with rich innate knowledge, children still must assign sound sequences to appropriate grammatical categories while determining the syntactic relations between these categories in their native language. Recently, a wealth of compelling experimental evidence has accumulated, suggesting that children do not initially use abstract linguistic categories. Instead, they seem to employ words at first as concrete individuals (rather than instances of abstract kinds), thereby challenging the usefulness of hypothesized innate grammatical categories (Tomasello, 2000). Whether we grant the presence of extensive innate knowledge or not, it seems clear that other sources of information are necessary to solve the bootstrapping problem.

Language-external information, such as correlations between the environment and semantic categories, may contribute to language acquisition by supplying a "semantic bootstrapping" solution (Pinker, 1984). However, because children learn linguistic distinctions that have no semantic basis (e.g., gender in French: Karmiloff-Smith, 1979), semantics cannot be the only source of information involved in solving the bootstrapping problem. Other sources of language-external constraints include cultural learning, indicated by a child's imitation of linguistic forms in socially conventional contexts (Tomasello, Kruger & Ratner, 1993). For example, a child may perceive that the idiom "*John let the cat out of the bag*," used in the appropriate context, means that John has revealed some sort of secret, and not that he released a feline from captivity. Despite both of these important language-external sources, to break down the linguistic forms into relevant units, it appears that correlation and cultural learning must be coupled with language-internal information.

We do not challenge the important role that the two foregoing sources of information play in language acquisition. We would argue, however, that language-internal information is fundamental to bootstrapping the child into syntax. Because language-internal input is rich in potential cues to linguistic structure, we offer a requisite feature of this information for syntax acquisition: Cues may only be partially reliable individually, and a learner must integrate an array of these cues to solve the bootstrapping problem. For example, a learner could use the tendency for English nouns to be longer than verbs to conjecture that *bonobo* is a noun, but the same strategy would fail for *ingratiate*. Likewise, although speakers tend to pause at syntactic phrase boundaries in a sentence, pauses also occur elsewhere during normal language production. And although it is a good distributional bet that the definite article *the* will precede a noun, so might adjectives, such as *silly*. The child therefore needs to integrate a great diversity of probabilistic cues to language structure. Fortunately, as we review in the next section, there is now extensive evidence that multiple probabilistic cues are available in language-internal input, that children are sensitive to them, and that they facilitate learning through integration.

Bootstrapping through Multiple Language-Internal Cues

We explore three sources of language-internal cues: phonological, prosodic, and distributional. Phonological information includes stress, vowel quality, and duration, and may help distinguish grammatical function words (e.g., determiners, prepositions, and conjunctions) from content words (nouns, verbs, adjectives, and adverbs) in English (e.g., Cutler, 1993; Gleitman & Wanner, 1982; Monaghan, Chater & Christiansen, 2005; Monaghan, Christiansen & Chater, 2007; Morgan, Shi, & Allopenna, 1996; Shi, Morgan, & Allopenna, 1998). Phonological information may also help separate nouns and verbs (Monaghan, Chater, & Christiansen, 2005; Monaghan, Christiansen, & Chater, 2007; Onnis & Christiansen, 2008). For example, English disyllabic nouns tend to receive initial-syllable (trochaic) stress whereas disyllabic verbs tend to receive final-syllable (iambic) stress, and adults are sensitive to this distinction (Kelly, 1988). Acoustic analyses have also shown that disyllabic words that are noun–verb ambiguous and have the same stress placement can still be differentiated by syllable duration and amplitude cue differences (Sereno & Jongman, 1995). Even 3-year-old children are sensitive to this stress cue, despite the fact that few multisyllabic verbs occur in child-directed speech (Cassidy & Kelly, 1991, 2001). Additional noun/verb cues in English likely include differences in word duration, consonant voicing, and vowel types, and many of these cues may be cross-linguistically relevant (see Kelly, 1992; Monaghan & Christiansen, 2008, for reviews).

Prosodic cues help word and phrasal/clausal segmentation and may reveal syntactic structure (e.g., Gerken, Jusczyk & Mandel, 1994; Gleitman & Wanner, 1982; Kemler-Nelson, Hirsh-Pasek, Jusczyk, & Wright Cassidy, 1989; Morgan, 1996). Acoustic analyses find that pause length, vowel duration, and pitch all mark phrasal boundaries in English and Japanese child-directed speech (Fisher & Tokura, 1996). Perhaps from utero (Mehler et al., 1988) and beyond, infants seem highly sensitive to such language-specific prosodic patterns (Gerken et al., 1994; Kemler-Nelson et al., 1989; for reviews, see Gerken, 1996; Jusczyk & Kemler-Nelson, 1996; Morgan, 1996). Prosodic information also improves sentence comprehension in 2-year-olds (Shady & Gerken, 1999). In experiments using adult participants, artificial language learning is facilitated in the presence of prosodic marking of syntactic phrase boundaries (Morgan, Meier & Newport, 1987; Valian & Levitt, 1996). Neurophysiological evidence in the form of event-related brainwave potentials (ERP) in adults shows that prosodic information has an immediate effect on syntactic processing (Steinhauer, Alter, & Friederici, 1999), suggesting a rapid, on-line role for this important cue. While prosody is influenced to some extent by a number of nonsyntactic factors, such as breathing patterns, resulting in an imperfect mapping between prosody and syntax (Fernald & McRoberts, 1996), infants' sensitivity to prosody argues for its likely contribution to syntax acquisition (Fisher & Tokura, 1996; Gerken 1996; Morgan, 1996).

Distributional characteristics of linguistic fragments at or below the word level may also provide cues to grammatical category. Morphological patterns across words may be informative—e.g., English words that are observed to have both *–ed* and *–s* endings are likely to be verbs (Maratsos & Chalkley, 1980). In artificial language learning experiments, adults acquire grammatical categories more effectively when they are cued by such word-internal patterns

(Brooks, Braine, Catalano & Brody, 1993; Frigo & McDonald, 1998). Corpus analyses reveal that word co-occurrence also gives useful cues to grammatical categories in child-directed speech (e.g., Mintz, 2003; Monaghan et al., 2005, 2007; Redington, Chater, & Finch, 1998). Given that function words primarily occur at phrase boundaries (e.g., initially in English and French and finally in Japanese), they can also help the learner by signaling syntactic structure. This idea has received support from corpus analyses (Mintz, Newport & Bever, 2002) and artificial language learning studies (Green, 1979; Morgan et al., 1987; Valian & Coulson, 1988). Finally, artificial language learning experiments indicate that duplication of morphological patterns across related items in a phrase (e.g., Spanish: *L*<u>os</u> *Estad*<u>os</u> *Unid*<u>os</u>) <COMP: Keep underline for clarity.> facilitates learning (Meier & Bower, 1986; Morgan et al., 1987).

It is important to note that there is ample evidence that children are sensitive to these multiple sources of information. After just 1 year of language exposure, the perceptual attunement of children likely allows them to make use of language-internal probabilistic cues (for reviews, see Jusczyk, 1997, 1999; Kuhl, 1999; Pallier, Christophe & Mehler, 1997; Werker & Tees, 1999). Through early learning experiences, infants already appear sensitive to the acoustic differences between function and content words (Shi, Werker & Morgan, 1999) and the relationship between function words and prosody in speech (Shafer, D. W. Shucard, J. L. Shucard & Gerken, 1998). Young infants are able to detect differences in syllable number among isolated words (Bijeljac, Bertoncini & Mehler, 1993). In addition, infants exhibit rapid distributional learning (e.g., Gómez & Gerken, 1999; Saffran, Aslin, & Newport, 1996; see Gómez & Gerken, 2000; Saffran, 2003 for reviews), and importantly, they are capable of multiple-cue integration (Mattys, Jusczyk, Luce, & Morgan, 1999; Morgan & Saffran, 1995). When facing the bootstrapping problem, children probably also benefit from characteristics of child-directed speech, such as the predominance of short sentences (Newport, Gleitman & Gleitman, 1977) and exaggerated prosody (Kuhl et al., 1997).

In summary, phonological information helps to distinguish function words from content words and nouns from verbs. Prosodic information helps word and phrasal/clausal segmentation, thus serving to uncover syntactic structure. Distributional characteristics aid in labeling and segmentation, and may provide further cueing of syntactic relations. Despite the value of each source, none of these cues in isolation suffices to solve the bootstrapping problem. The learner must integrate these multiple cues to overcome the limited reliability of each individually. This review has indicated that a range of language-internal cues is available for language acquisition, that these cues affect learning and processing, and that mechanisms exist for multiple-cue integration. What is yet unknown is how far these cues can be combined to solve the bootstrapping problem (Fernald & McRoberts, 1996). Here we present connectionist simulations to demonstrate that efficient and robust computational mechanisms exist for multiple-cue integration (see also the chapters in this volume by Hannon, Kirkham, and Saffran, for evidence from human infant learning).

SIMULATION 1: MULTIPLE-CUE INTEGRATION

Although the multiple-cue approach is gaining support in developmental psycholinguistics, its computational efficacy still remains to be established. The simulations reported in this chapter are therefore intended as a first step toward a computational approach to multiple-cue integration, seeking to test its potential value in syntax acquisition. Based on our previous experience with modeling multiple-cue integration in speech segmentation (Christiansen, Allen, & Seidenberg, 1998), we used a simple recurrent network (SRN; Elman, 1990) to model the integration of multiple cues. The SRN is feed-forward neural network equipped with an additional copy-back loop that permits the learning and processing of temporal regularities in the stimuli presented to it (see Figure 5.1). This makes it particularly suitable for exploring the acquisition of syntax, an inherently temporal phenomenon.

INSERT FIGURE 5.1 ABOUT HERE

The networks were trained on corpora of artificial child-directed speech generated by a grammar that includes three probabilistic cues to grammatical structure: word length, lexical stress, and pitch. The grammar (described further below) was motivated by considering frequent constructions in child-directed speech in the CHILDES database (MacWhinney, 2000). Simulation 1 demonstrates how the integration of these three cues benefits the acquisition of syntactic structure by comparing performance across the eight possible cue combinations ranging from the absence of cues to the presence of all three.

Method

Networks

Ten networks were trained per condition, with an initial randomization of network connections in the interval [–0.1, 0.1]. Learning rate was set to 0.1, and momentum to 0. Each input to the networks contained a localist representation of a word (one unit = one word) and a set of cue units depending on cue condition. Words were presented one by one, and networks were required to predict the next word in a sentence along with the corresponding cues for that word. With a total of 44 words (see below) and a pause marking boundaries between utterances, the networks had 45 input units. Networks in the condition with all available cues had an additional five input units. The number of input and output units thus varied between 45 and 50 across conditions. Each network had 80 hidden units and 80 context units.

Materials

We constructed an idealized but relatively complex grammar based on independent analyses of child-directed speech corpora (Bernstein-Ratner, 1984; Korman, 1984) and a study of child-directed speech by mother–daughter pairs (Fisher & Tokura, 1996). As illustrated in Table 5.1, the grammar included three primary sentence types: declarative, imperative, and interrogative sentences. Each type consisted of a variety of common utterances reflecting the child's exposure. For example, declarative sentences most frequently appeared as transitive or intransitive verb constructions (*the boy chases the cat*, *the boy swims*), but also included predication using *be* (*the horse is pretty*) and second person pronominal constructions commonly found in child-directed corpora (*you are a boy*). Interrogative sentences were composed of wh-questions (*where are the boys?*, *where do the boys swim?*), and questions formed by using auxiliary verbs (*do the boys walk?*, *are the cats pretty?*). Imperatives were the simplest class of sentences, appearing as intransitive or transitive verb phrases (*kiss the bunny*, *sleep*). Subject–verb agreement was upheld in the grammar, along with appropriate determiners accompanying nouns (*the cars* vs. **a cars*).

Each word was assigned a unit for input into the model, and we added a number of units to represent cues. Two basic cues were available to all networks. The fundamental distributional information inherent in the grammar could be exploited by all networks in this simulation. As a second basic cue, utterance-boundary pauses signaled grammatically distinct utterances with 92% reliability (Broen, 1972). This was encoded as a single unit that was activated at the end of all but 8% of the sentences. Other semireliable prosodic and phonological cues accompanied the phrase-structure grammar: word length, stress, and pitch. Network groups were constructed using different combinations of these three cues. Cassidy and Kelly (1991) demonstrated that syllable count is a cue available to English speakers to distinguish nouns and verbs. They found that the probability of a single syllable word to be a noun rather than a verb is 38%. This probability rises to 76% at two syllables, and 92% at three. We selected verb and noun tokens that exhibited this distinction, whereas the length of the remaining words was typical for their class (i.e., function words tended to be monosyllabic). Word length was represented in terms of three units using thermometer encoding—that is, one unit would be on for monosyllabic words, two for bisyllabic words, and three for trisyllabic words. Pitch change is a cue associated with syllables that precede pauses. Fisher and Tokura (1996) found that these pauses signaled grammatically distinct utterances with 96% accuracy in child-directed speech, allowing pitch to serve as a cue to grammatical structure. In the networks, this cue was a single unit that would be activated at the final word in an utterance. Finally, we used a single unit to encode lexical stress as a possible cue to distinguish stressed content words from the reduced, unstressed form of function words. This unit would be on for all content words.

INSERT TABLE 5.1 ABOUT HERE

Procedure

Eight groups of networks, one for each combination of cues (all cues, 2 cues, 1 cue, or none), were trained on corpora consisting of 10,000 sentences generated from the grammar. Each network within a group was trained on a different randomized training corpus. Training consisted of 200,000 input/output presentations (words), or approximately 5 passes through the training corpus. Each group of networks had cues added to its training corpus depending on cue condition. Networks were expected to predict the next word in a sentence, along with the appropriate cue values. A corpus consisting of 1,000 novel sentences was generated for testing. Performance was measured by assessing the networks' ability to predict the next set of grammatical items given prior context. Importantly, this measure did not include predictions of cue information, and all network conditions were thus evaluated by exactly the same performance criterion.

To provide a statistical benchmark with which to compare network performance, we trained bigram and trigram models on the same corpora as the networks. These finite-state models, borrowed from computational linguistics, provide a simple prediction method based on strings of two (bigrams) or three (trigrams) consecutive words. Comparisons with these simple models provide an indication of whether the networks are learning more than simple two- or three-word associations.

Results

After training, SRNs trained with localist output representations will produce a distributional pattern of activation closely corresponding to a probability distribution of possible next items. In order to assess the overall performance of the SRNs, we made comparisons between network output probabilities and the full conditional probabilities given the prior context. For example, the full conditional probabilities given the context of "*The boy chases*..." can be represented as a vector containing the probabilities of being the next item in this sentence for each of the 44 words in the vocabulary and the pause. To ensure that our performance measure can deal with novel test sentences not seen during training, we estimate the prior conditional probabilities based on lexical categories rather than individual words (Christiansen & Chater, 1999). Suppose, in the example above, that every continuation of this sentence fragment in the training corpus always involved the indefinite determiner "*a*" (as in "*The boy chases a cat*"). If we did not base our full conditional probability estimates on lexical categories, we would not be able to assess SRN performance on novel sentences in which the definite determiner *"the"* followed the example fragment (as in "*The boy chases the cat*"). Formally, we thus have the following Equation 1 with c_i denoting the category of the *i*th word in the sentence:

$$P(c_p \mid c_1, c_2, \ldots, c_{p-1}) \cong \frac{\text{Freq}(c_1, c_2, \ldots, c_{p-1}, c_p)}{\text{Freq}(c_1, c_2, \ldots, c_{p-1})} \qquad (5.1)$$

where the probability of getting some member of a given lexical category as the *p*th item, c_p, in a sentence is conditional on the previous p–1 lexical categories. Note that for the purpose of performance assessment, singular and plural nouns are assigned to separate lexical categories throughout Simulations 1–4, as are singular and plural verbs. Given that the choice of lexical items for each category is independent, and that each word in a category is equally frequent, the probability of encountering a particular word w_n, which is a member of a category c_p, is simply inversely proportional to the number of items, C_p, in that category. So, overall, we have the following equation:

$$P(w_n \mid c_1, c_2, \ldots, c_{p-1}) \cong \frac{\text{Freq}(c_1, c_2, \ldots, c_{p-1}, c_p)}{\text{Freq}(c_1, c_2, \ldots, c_{p-1}) C_p} \qquad (5.2)$$

If the networks are performing optimally, then the vector of output unit activations should exactly match these probabilities. We evaluate the degree to which each network performs successfully by measuring the mean squared error between the vectors representing the network's output and the conditional probabilities (with 0 indicating optimal performance).

All networks achieved better performance than the standard bigram/trigram models (p-values < .0001), suggesting that the networks had acquired knowledge of syntactic structure beyond the information associated with simple pairs or triples of words. Figure 5.2A illustrates the best performance achieved by the trigram model as well as SRNs provided with no cues (the baseline network), a single cue (length, stress, or prosody), and three cues. The nets provided with one or more phonological/prosodic cues achieved significantly better performance than baseline networks (p-values < .02). Using trigram performance as criterion, all multiple-cue networks surpassed this level of performance faster than the baseline networks as shown in Figure 5.2B (p-values < .002). Moreover, the three-cue networks were significantly faster than the single-cue networks (p-values < .001). Finally, using Brown-Forsyth tests for variability in the final level of performance, we found that the three-cue networks also exhibited significantly more uniform learning than the baseline networks ($F(1,18) = 5.14$, $p < .04$), as depicted in Figure 5.2C.

INSERT FIGURE 5.2 ABOUT HERE

SIMULATION 2: SENTENCE COMPREHENSION IN 2-YEAR-OLDS

Simulation 1 provides evidence for the general feasibility of multiple-cue integration for supporting syntax learning. To further demonstrate the relevance of the model to language development, closer contact with human data is needed (Christiansen & Chater, 2001). In the current simulation, we demonstrate that the three-cue networks from Simulation 1 are able to accommodate experimental data showing that 2-year-olds can integrate grammatical markers (function words) and prosodic cues in sentence comprehension (Shady & Gerken, 1999: Experiment 1). In this study, children heard sentences, such as (1) [see below], in one of three prosodic conditions depending on pause location: early natural [e], late natural [l], and unnatural [u]. Each sentence moreover involved one of three grammatical markers: grammatical (*the*), ungrammatical (*was*), and nonsense (*gub*).

1. Find [e] the/was/gub [u] dog [l] for me.

The child's task was to identify the correct picture corresponding to the target noun (*dog*). Children performed the task best when the pause location delimited a phrasal boundary (early/late), and with the grammatical marker *the*. Simulation 2 models these data by using comparable stimuli and assessing noun unit activations.

Method

Networks

Twelve three-cue networks of the same architecture and training used in Simulation 1 were used in each prosodic condition in the infant experiment. This number was chosen to match the number of infants in the Shady and Gerken (1999) experiment. An additional unit was added to the networks to encode the nonsense word (*gub*) in Shady and Gerken's experiment.

Materials

We constructed a sample set of sentences from our grammar that could be modified to match the stimuli in Shady and Gerken. Twelve sentences for each prosody condition (pause location) were constructed. Pauses were simulated by activating the utterance-boundary unit. Because these pauses probabilistically signal grammatically of distinct utterances, the utterance-boundary unit provides an approximation of what the children in the experiment would experience. Finally, the nonsense word was added to the stimuli for the within group condition (grammatical vs. ungrammatical vs. nonsense). Adjusting for vocabulary differences, the networks were tested on comparable sentences, such as (2):

2. Where does [e] the/is/gub [u] dog [l] eat?

Procedure

Each group of networks was exposed to the set of sentences corresponding to its assigned pause location (early vs. late vs. unnatural). No learning took place, since the fully trained networks were used. To approximate the picture selection task in the experiment, we measured the degree to which the networks would activate the groups of nouns following *the*/*is*/*gub*. The two conditions were expected to affect the activation of the nouns.

Results

The human results for the prosody condition in Shady and Gerken (1999) is depicted in Figure 5.3A. They reported a significant effect of prosody on the picture selection task. The same was true for our networks ($F(2,33) = 1{,}253.07$, $p < .0001$), and the pattern of noun activations closely resembles that of the toddlers' correct picture choice as evidenced by Figure 5.3B. The late natural condition elicited the highest noun activation, followed by the early natural condition, and with the unnatural condition yielding the least activation. The experiment also revealed an effect of grammaticality as can be seen from the human data shown in Figure 5.3C. We similarly obtained a significant grammaticality effect for our networks ($F(2,70) = 69.85$, $p < .0001$), which, as illustrated by Figure 5.3D, produced the highest noun activation following the determiner, followed by the nonsense word, and lastly for the ungrammatical word. Again, the network results match the pattern observed for the toddlers. One slight discrepancy is that the networks are producing higher noun activation following the nonsense word compared to the ungrammatical marker. This result is however consistent with the results from a more sensitive picture selection task, showing that children were more likely to end up with a semantic representation of the target following nonsense syllables compared to incorrectly used morphemes (Carter & Gerken, 1996). Thus, the results suggest that the syntactic knowledge acquired by the networks mirrors the kind of sensitivity to syntactic relations and prosodic content observed in human children. Together with Simulation 1, the results also demonstrate that multiple-cue integration may both facilitate syntax acquisition, and underlie some patterns of linguistic skill observed early on in human performance. In the next simulation, we show that the multiple-cue perspective can simulate possible prosodic scaffolding that occurs much earlier in development: prenatal attunement to prosody.

INSERT FIGURE 5.3 ABOUT HERE

SIMULATION 3: THE ROLE OF PRENATAL EXPOSURE

Studies of 4-day-old infants suggest that the attunement to prosodic information may begin prior to birth (Mehler et al., 1988). We suggest that this prenatal exposure to language may provide a scaffolding for later syntactic acquisition by initially focusing learning on certain aspects of prosody and gross-level properties of phonology (such as word length) that later will play an important role in postnatal multiple-cue integration. In the current simulation, we test this hypothesis using the connectionist model from Simulations 1 and 2. If this scaffolding hypothesis is correct, we would expect that prenatal exposure corresponding to what infants receive in the womb would result in improved acquisition of syntactic structure.

Method

Networks

Ten SRNs were used in both prenatal and nonprenatal groups, with the same initial conditions and training details as Simulation 1. Each network was supplied with the full range of cues used in Simulation 1.

Materials

A set of “filtered” prenatal stimuli was generated using the same grammar as previously (Table 5.1), with the exception that input/output patterns now ignored individual words and only involved the units encoding word length, stress, pitch change and utterance boundaries. The postnatal stimuli were the same as in Simulation 1.

Procedure

The networks in the prenatal group were first trained on 100,000 input/output filtered presentations drawn from a corpus of 10,000 new sentences. Following this prenatal exposure, the nets were then trained on the full input patterns exactly as in Simulation 1. The nonprenatal group only received training on the postnatal corpora. As previously, networks were required to predict the following word and corresponding cues. Performance was again measured by the prediction of following words, ignoring the cue units.

Results

Both network groups exhibited significantly higher performance than the bigram/trigram models ($F(1,18) = 25.32$, $p < .0001$ for prenatal, $F(1,18) = 12.03$, $p < .01$ for nonprenatal), again indicating that the networks are acquiring complex grammatical regularities that go beyond simple adjacency relations. We compared the performance of the two network groups across different degrees of training using a two-way analysis of variance with training condition (prenatal vs. nonprenatal) as the between-network factor and amount of training as within-network factor (five levels of training measured in 20,000 input/output presentation intervals). There was a main effect of training condition ($F(1,18) = 12.36$, $p < .01$), suggesting that prenatal exposure significantly improved learning. A main effect of degrees of training ($F(9,162) = 15.96$, $p < .001$) reveals that both network groups benefited significantly from training. An interaction between training conditions and degrees of training indicates that the prenatal networks learned significantly better than postnatal networks ($F(1,18) = 9.90$, $p < .01$). Finally, as illustrated by Figure 5.4, prenatal input also resulted in faster learning (measured in terms of the amount of training needed to surpass the trigram model; $F(1,18) = 9.90$, $p < .01$). The exposure to prenatal input—void of any information about individual words—promotes better performance on the prediction task as well as faster learning overall. This provides computational support for the prenatal scaffolding hypothesis, derived as a prediction from the multiple-cue perspective on syntax acquisition.

INSERT FIGURE 5.4 ABOUT HERE

SIMULATION 4: MULTIPLE-CUE INTEGRATION WITH USEFUL AND DISTRACTING CUES

So far, simulations have demonstrated the importance of cue integration in syntax acquisition, that integration can match data obtained in infant experiments, and that this perspective can provide novel predictions in language development. A possible objection to these simulations is that our networks succeed at multiple-cue integration because they are only provided with cues that are at least partially relevant for syntax acquisition. Consequently, performance may potentially drop significantly if the networks themselves had to discover which cues were partially relevant and which are not. Simulation 4 therefore tests the robustness of our multiple-cue approach when faced with additional, uncorrelated distractor cues. Accordingly, we added three distractor cues to the previous three reliable cues. These new cues encoded the presence of word-initial vowels, word-final voicing, and relative (male/female) speaker pitch—all acoustically salient in speech, but which do not appear to cue syntactic structure.

Method

Networks

Networks, groups, and training details were the same as in Simulation 3, except for three additional input units encoding the distractor cues.

Materials

The three distractor cues were added to the stimuli used in Simulation 3. Two of the cues were phonetic and therefore available only in postnatal training. The word-initial vowel cue appears in all words across classes. The second distractor cue, word-final voicing, also does not provide useful distinguishing properties of word classes. Finally, as an additional prenatal and postnatal cue, overall pitch quality was added to the stimuli. This was intended to capture whether the speaker was female or male. In prenatal training, this probability was set to be extremely high (90%), and lower in postnatal training (60%). In the womb, the mother's voice naturally provides most of the input during the final trimester when the infant's auditory system has begun to function (Rubel, 1985). The probability used here was intended to capture the likelihood that some experience would derive from other speakers as well. In postnatal training, this probability drops, representing exposure to male members of the linguistic community, but still favoring mother–child interactions.

Procedure

Prenatal stimuli included the three previous semireliable cues, and only the additional prosodic, distractor cue encoding relative speaker pitch. In the postnatal stimuli, all three distractor cues were added. Training and testing details were the same as in Simulation 3.

Results

As in Simulations 1 and 3, both groups performed significantly better than the bigram/trigram models ($F(1,18) = 18.95$, $p < .0001$ for prenatal, and $F(1,18) = 14.27$, $p < .001$ for nonprenatal). We repeated the two-factor analysis of variance computed for Simulation 2, revealing a main effect for training condition ($F(1,18) = 4.76$, $p < .05$) and degrees of training ($F(9,162) = 13.88$, $p < .0001$). This indicates that the presence of the distractor cues did not hinder the improved performance following prenatal language exposure. As in Simulation 3, the prenatal networks learned comparatively faster than the nonprenatal networks ($F(1,18) = 5.31$, $p < .05$). To determine how the distractor cues affected performance, we compared the prenatal condition in Simulation 3 with that of the current simulation. There was no significant difference in performance across the two simulations ($F(1,18) = 0.13$, $p = .72$). Moreover, as shown in Figure 5.5, there was no difference in the speed of learning between the SRNs trained only with good cues and those whose input included distractor cues ($F(1,18) = .57$, $p = .46$). A further comparison between these nonprenatal networks and the bare networks in Simulation 1 showed that the networks trained with cues of mixed reliability significantly outperformed networks trained without any cues ($F(1,18) = 14.27$, $p < .001$). This indicates that the uncorrelated cues did not prevent the networks from integrating the partially reliable ones toward learning grammatical structure. Together with the first three simulations, Simulation 4 demonstrates that SRNs can integrate multiple cues efficiently when exposed to relatively complex artificial corpora. Next, we scale up the model to deal with naturalistic child-directed speech.

INSERT FIGURE 5.5 ABOUT HERE

SIMULATION 5: MULTIPLE-CUE INTEGRATION WITH FULL-BLOWN CHILD-DIRECTED SPEECH

In this final simulation, we take a further step toward describing the computational underpinnings of multiple-cue integration. The previous series of simulations have demonstrated that SRNs provide a suitable model for integrating multiple cues when exposed to input generated by a psychologically motivated artificial grammar. Here we further show that the SRN scales up to deal with real child-directed speech. In particular, we seek to determine the extent to which these networks are sensitive to the lexical category information present in the set of phonological cues. To accomplish this task, we set up two identical groups of networks, each provided with a different encoding of the corpus. The encoding of the first corpus was based on 16 phonological cues, previously shown by Monaghan et al. (2005) to provide information useful for syntax acquisition. The second set of input was encoded using the same cue vectors but randomized across lexical categories. Possible performance differences in networks trained with these different input sets would be due to lexical category information revealed by the multiple phonological cues.

Method

Networks

Ten SRNs were used for the phonetic-input condition and the random-input condition, with an initial weight randomization in the interval [–0.1, 0.1]. A different random seed was used for each simulation. Learning rate was set to 0.1 and momentum to 0.7. Each input to the network contained a thermometer encoding for each of the 16 phonological cues from Monaghan et al. (2005), listed in Table 5.2. This encoding required 43 units (each of them in a range from 0 to 1) and a pause marking boundaries between utterances, resulting in the networks having 44 input units. Each output was encoded using a localist representation consisting of 14 different lexical categories and a pause marking boundaries between utterances, resulting in networks with 15 output units. Each network furthermore was equipped with 88 hidden units and 88 context units.

INSERT TABLE 5.2 ABOUT HERE

Materials

We trained and tested the network on a corpus of child-directed speech (Bernstein-Ratner, 1984). This corpus contains speech recorded from nine mothers speaking to their children over a 4- to 5-month period when the children were between the ages of 1 year and 1 month to 1 year and 9 months. The corpus includes 1,371 word types and 33,035 tokens distributed over 10,082 utterances. The sentences incorporate a number of different types of grammatical structures, showing the varied nature of the linguistic input to children. Utterances range from declarative sentences (*Oh you need some space*) to wh-questions (*Where's my apple*) to one-word utterances ("*Uh*" or "*hello*"). Each word in the corpus corresponded to one of the 14 following lexical categories: nouns (19.5%), verbs (18.5%), adjectives (4%), numerals (<0.1%), adverbs (6.5%), articles (6.5%), pronouns (18.5%), prepositions (5%), conjunctions (4%), interjections (7%), complex contractions (8%), abbreviations (<0.1%), infinitive markers (1.2%), and proper names (1.2%). The training set consisted of 9,072 sentences (29,930 word tokens) from the original corpus. A separate test set consisted of 963 additional sentences (2,930 word tokens).

Each word was encoded in terms of the following 16 phonological cues from Table 5.2: number of phonemes (1–11), number of syllables (1–5), stress position (0 = no stress, 1 = 1st syllable stressed, etc.), proportion of reduced vowels (0–1), proportion of coronal consonants (0–1), number of consonants in onset (1–3), consonant complexity (0–1), initial /D/ (1 if begins /D/, 0 otherwise), reduced first vowel (1 if first vowel is reduced, 0 otherwise), any stress (0 if no stress, 1 otherwise), final inflection (0 if none, /@d/ or /Id/, 1 if present), stress vowel position (from front to back, 1–3), vowel position (mean position of vowels, from front to back, 1–3), final consonant voicing (0: vowel, 1: voiced, 2: unvoiced), proportion of nasal consonants (0–1) and mean height of vowels (0–3). The cues that assume only binary values were encoded using a single unit (e.g., "any stress", "initial /D/"). The cues that take on values between 0 and 1 (e.g., proportion of vowel consonants) were also encoded using a single unit with a decimal number, whereas the cues that assume values in a broader range (e.g., number of syllables) were represented using a thermometer encoding; for example, one unit would be on for monosyllabic words, two for bisyllabic words, and so on. Finally we used a single unit that would be activated at pauses between utterances.

The random-input networks were trained using input for which we randomly distributed the multiple-cue vectors among all the words in the corpus. Thus, the vector encoding for a given word would be randomly reassigned to a different word in the corpus regardless of its lexical category. Each phonological vector was assigned to only one word. Moreover, each token of a word was represented using the same random vector for all occurrences of that word in the test and training sets.

Procedure

Ten networks were trained on phonological cues and 10 control networks were trained on the random vectors. Training consisted of one pass through the training corpus. We used the same 10 random seeds for both simulation conditions. The networks were trained to predict the lexical category of the next word. The task of mapping phonological cues onto lexical categories may seem somewhat artificial because children are not provided directly with the lexical categories of the words to which they are exposed. However, children do learn early on to use pragmatic and other cues to discover the meaning of words. Given that the networks in our simulations only have access to linguistic information, we see lexical categories as a “stand-in” for more ecologically valid cues that we hope to be able to include in future work.

Results

We recorded the output vectors for the two groups of networks. Because the output consisted of localist representations for each lexical category (one unit = one lexical category) along with the utterance-final pause, we could use Equation 5.1 to estimate the full conditional probabilities, comparing network predictions to the full conditional probabilities for the next lexical category using the mean cosine of the angle between the two vectors (with 1 corresponding to optimal performance). We compared the predictions of the phonetic-input networks with those of the random-input networks. Figure 5.6A shows a comparison of test–set performance for the phonetic-input networks with that of the random-input networks. The phonetic-input networks were significantly better than the random-input networks at predicting the next combination of lexical categories (p-values < .00005). These results suggest that distributional information is generally a stronger cue than phonological information, even though the latter does lead to better learning overall. However, phonological information may provide the networks with a better basis for processing novel lexical items. Next, we probe the internal representations of the two sets of networks in order to gain further insight into their performance differences.

INSERT FIGURE 5.6 ABOUT HERE

Probing the Internal Representations

Simulation 5 indicated that the phonetic-input networks did not benefit as much as one perhaps would have expected from the information provided by the phonological cues. However, the networks may nonetheless use this information to develop internal representations that better encode differences between lexical categories. This may allow them to go beyond the phonetic input and integrate it with the distributional information derived from the sequential order in which these vectors were presented. To investigate these possibilities, we carried out a series of discriminant analyses of network hidden unit activations as well as of the phonetic input vectors, focusing on the representations of nouns and verbs.

Method

Informally, a linear discriminant analysis allows us to determine the degree to which it is possible to separate a set of vectors into two (or more) groups based on the information contained in those vectors. In effect, we attempt to use a linear plane to split the hidden unit space into a group of noun vectors and a group of verb vectors. Using discriminant analyses, we can statistically estimate the degree to which this split can be accomplished given a set of vectors.

We recorded the hidden unit activations from the two sets of networks in Simulation 5. The hidden unit activations were recorded for 200 novel nouns and 200 novel verbs occurring in unique sentences taken from other CHILDES corpora (MacWhinney, 2000). The hidden unit activations were labeled such that each corresponded to the particular lexical category of the input presented to the network (though the networks did not receive this information as input). For example, a vector would be labeled a noun vector when the hidden unit activations were recorded for a noun (phonetic) input vector. We also included a condition in which the noun/verb labels were randomized with respect to the hidden unit vectors for both sets of networks, in order to establish a random control.

Results

We first compared the categorization performance of the two sets of networks, as illustrated in Figure 5.6B. The phonetic-input networks had developed hidden unit representations that allowed them to correctly separate 80.30% of the 400 nouns and verbs. This was significantly better than the random-input networks, which only achieved 73.15% correct separation ($t(8) = 5.89$, $p < .0001$). Both sets of networks surpassed their respective randomized controls (phonetic-input control: 69.05% – $t(8) = 11.51$, $p < .0001$; random-input control: 68.20% – $t(8) = 3.92$, $p < .004$). The controls for the two sets of networks were not significantly different from each other ($t(8) = 0.82$, $p > .43$). As indicated by our previous analyses of phonetic cue information in child-directed speech (Monaghan et al., 2005), the phonetic input vectors contained a considerable amount of information about lexical categories, allowing for 67.25% correct separation of nouns and verbs, but still significantly below the performance of the phonetic-input networks ($t(4) = 25.97$, $p < .0001$). The random-input networks also surpassed the level of separation afforded by their input vectors (59.00% – $t(4) = 12.80$, $p < .0001$).

The results of the hidden-unit discriminant analyses suggest that not only did the phonetic-input networks develop internal representations better suited for distinguishing between nouns and verbs, but they also went beyond the information afforded by the phonetic input and integrate it with distributional information. Crucially, the phonetic-input vectors were able to surpass the random-input networks, despite that the latter was also able to use distributional information to go beyond the input. Consistent phonological information thus appears to be important for network generalization to novel nouns and verbs.

GENERAL DISCUSSION

As described in an earlier part of this chapter, children who are learning syntax face a complex "chicken-and-egg" bootstrapping problem. A growing bulk of evidence from developmental cognitive science has suggested that a solution may come from a process of integrating multiple sources of probabilistic information, each of which is individually unreliable, but jointly advantageous (cf. Smith & Pereia chapter in this volume). What has so far been lacking is a demonstration of the computational feasibility of this approach and the series of simulations reported here takes a first step toward accomplishing this. We have demonstrated that providing SRNs with prosodic and phonological cues significantly improves their acquisition of syntactic structure (Simulation 1), and that the three-cue networks can mimic children's sensitivity to both prosodic and grammatical cues in sentence comprehension (Simulation 2). The model illustrates the potential value of prenatal exposure (Simulation 3) and provides evidence for the robustness of multiple-cue integration, since highly unreliable cues did not interfere with the integration process (Simulation 4). Finally, we expanded these results by showing that SRNs can also utilize highly probabilistic information found in 16 phonological cues in the service of syntactic acquisition when trained on a naturalistic corpus of child-directed speech (Simulation 5). Analysis of the networks' hidden unit activations provided further evidence that the integration of phonological and distributional cues during learning leads to more robust internal representations of lexical categories, at least when it comes to distinguishing between the two major categories of nouns and verbs.

Overall, the simulation results presented in this chapter provide support not only for the multiple-cue integration approach in general, but also for using neural network architectures to explore the integration of distributional, prosodic, and phonological information in language acquisition. Some researchers have challenged the value of multiple probabilistic cues (e.g., Fernald & McRoberts, 1996), but we have computationally demonstrated that their integration results in faster, better, and more uniform learning, even in the face of distracting information. Our simulations, along with artificial language learning experiments (Billman, 1989; Brooks et al., 1993; McDonald & Plauche, 1995; Morgan et al., 1987), underscore multiple-cue integration as a means of facilitating the complex task of syntax acquisition.

We have elsewhere explored the evolutionary emergence of phonological cues in agent-based simulations (Christiansen & Dale, 2004). In these evolutionary simulations, languages were mutated slightly across generations of randomized SRN learners. For any given generation, the languages best learned by the networks were allowed to be passed down to the next generation. Results showed that there emerges cross-linguistic variation in stable linguistic cues. Nevertheless, observed stable cue systems were consistent in that syntactic categories were marked by phonological cues, as found in English, French, Japanese, and other languages (as reviewed above). This stability was particularly strong when languages had larger lexicons, indicating that multiple-cue integration may have contributed to language evolution by aiding a learner's acquisition of growing set of lexical items and classes.

Because different natural languages employ different constellations of cues to signal syntactic distinctions, an important question for further research is exactly how a child's learning mechanisms discover which cues are relevant and for which aspects of syntax. This problem is compounded by the fact that the same cue may work in different directions across different languages. A case in point is that nouns tend to contain more vowels and fewer consonants than verbs in English, whereas nouns and verbs in French show the opposite pattern (Monaghan et al., 2007). So how can the child learn which cues are relevant and in which direction? One possibility may be to encode the correlations between cues in the linguistic environment. This view is supported by related mathematical analyses based on the Vapnik-Chervonenkis (VC) dimension (Abu-Mostafa, 1993), showing that the integration of multiple "hints" or cues of correlated information reduces the number of hypotheses a learning system has to entertain. The VC dimension specifies an upper bound for the amount of input needed by a learning process that starts with a set of hypotheses about a task solution. Cue information may lead to a reduction in the VC dimension by weeding out unhelpful hypotheses and thus lowering the number of examples needed to find a solution. In other words, the integration of multiple cues may reduce learning time by reducing the number of steps necessary to find an appropriate function approximation, as well as reduce the set of candidate functions considered, thus potentially ensuring better generalization.

More generally, the development of computational multiple-cue integration models is still in its infancy. There now exists a wealth of support for the usefulness of multiple probabilistic cues for language acquisition, and although theoretical models abound (e.g., Gleitman & Wanner, 1982; and contributions in Morgan & Demuth, 1996; Weissenborn & Höhle, 2001), only a few psychologically plausible computational models for multiple-cue integration are on offer (e.g., Cartwright & Brent, 1997). Extant models tend to capture the end-state of learning rather the developmental process itself. This approach cannot identify the time course of different cues as they become important for acquisition. For example, the ability to use visual context information to resolve a syntactically ambiguous sentence does not appear until about 8 years of age, considerably later than the knowledge of constraints on constructions that may follow specific verbs (Snedeker & Trueswell, 2004). To reveal cue integration and its development, models must capture the developmental trajectory of cue use across different phases of language acquisition. We anticipate that the availability of so-called "dense" corpora, which sample the child's input at a higher frequency (e.g., Behrens, 2006; Maslen, Theakston, Lieven, & Tomasello, 2004), will help the development of such constructivist-oriented models of language acquisition.

Future work should therefore provide more detailed analysis of the developmental trajectory of multiple-cue integration. Most work on cue availability in the child's environment makes the simplifying assumption that all information is available to the child simultaneously. This is an oversimplification: Children's productions indicate that the whole of language is not acquired in one step, but that overlapping phases of acquisition occur, where learning progress at any one time relies on progress that preceded it. Attempts to explain and exploit these learning phases in computational models has been successful in accounting for early processing constraints that facilitate later learning of complex syntactic structures (Elman, 1993), phrasal productions and errors in young children (Freudenthal, Pine, & Gobet, 2005), and the development of the lexicon (Steyvers & Tenenbaum, 2005). Such approaches could equally be applied to the computational simulation of multiple-cue integration reported in this chapter: The reliability of phonological, prosodic, or distributional cues could be based on the most frequent, or earliest-learned words, and constructed incrementally, and such a constructivist approach would enhance the cognitive plausibility of the availability and process of use of such cues by the developing child.

The wide array of phonological, prosodic, and distributional information sources in primary linguistic input may make the child's learning task substantially easier than it might seem when we consider only the complexities of syntax that they acquire. A domain-general learning mechanism, such as the SRN architecture used here, can capitalize on this rich information to acquire deep domain-specific knowledge that emerges through developmental time. Along with this language-internal information, surely innate and language-external constraints also contribute to the task, and future work should aim to integrate all three fundamental sources of constraints. We have nevertheless shown that even with relatively simple domain-general assumptions about the learner, multiple-cue integration can facilitate the complex task of syntax acquisition. Theories of the language learner therefore should not overburden innate and language-external constraints where language-internal multiple-cue integration can help.

ACKNOWLEDGMENTS

This research was supported in part by a Human Frontiers Science Program Grant (RGP0177/2001-B) to M.H.C. Some of the material in this chapter was adapted from

Christiansen, M. H., & Dale, R. (2001), Integrating distributional, prosodic and phonological information in a connectionist model of language acquisition, in *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 220–225), Mahwah, NJ: Lawrence Erlbaum, and Reali, F., Christiansen, M. H., & Monaghan, P. (2003), Phonological and distributional cues in syntax acquisition: Scaling up the connectionist approach to multiple-cue integration, in *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 970–975), Mahwah, NJ: Lawrence Erlbaum.

References

Abu-Mostafa, Y. S. (1993) Hints and the VC dimension. *Neural Computation, 5*, 278–288.

Behrens, H. (2006). The input–output relationship in first language acquisition. *Language and Cognitive Processes, 21*, 2–24.

Bernstein-Ratner, N. (1984). Patterns of vowel modification in motherese. *Journal of Child Language*, *11*, 557–578.

Bijeljac, R., Bertoncini, J., & Mehler, J. (1993). How do 4-day-old infants categorize multisyllabic utterances? *Developmental Psychology*, *29*, 711–721.

Billman, D. (1989). Systems of correlations in rule and category learning: Use of structured input in learning syntactic categories. *Language and Cognitive Processes*, *4,* 127–155.

Broen, P. (1972). *The verbal environment of the language-learning child*. ASHA Monographs, No. 17. Washington, DC: American Speech and Hearing Society.

Brooks, P. J., Braine, M. D., Catalano, L. & Brody, R. E. (1993). Acquisition of gender-like noun subclasses in an artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language*, *32,* 76–95.

Carter, A. & Gerken, L. A. (1996). Children's use of grammatical morphemes in on-line sentence comprehension. In E. Clark (Ed.), *Proceedings of the Twenty-Eighth Annual Child Language Research Forum* (Vol. 29). Palo Alto, CA: Stanford University Press.

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

Cassidy, K. W., & Kelly, M. H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language*, *30*, 348–369.

Cassidy, K. W., & Kelly, M. H. (2001). Children's use of phonology to infer grammatical class in vocabulary learning. *Psychonomic Bulletin and Review*, *8*, 519–523.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, *13*, 221–268.

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science, 23,* 157–205.

Christiansen, M. H., & Chater, N. (2001). Connectionist psycholinguistics: Capturing the empirical data. *Trends in Cognitive Sciences*, *5,* 82–88.

Christiansen, M. H., & Dale, R. (2004). The role of learning and development in the evolution of language. A connectionist perspective. In D. Kimbrough Oller & U. Griebel (Eds.), *Evolution of communication systems: A comparative approach. The Vienna Series in Theoretical Biology* (pp. 90–109). Cambridge, MA: MIT Press.

Cutler, A. (1993). Phonological cues to open-and closed-class words in the processing of spoken sentences. *Journal of Psycholinguistic Research*, *22*, 109–131.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

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

Fernald, A., & McRoberts, G. (1996). Prosodic bootstrapping: A critical analysis of the argument and the evidence. In J. L. Morgan & K. Demuth (Eds.), *From Signal to syntax* (pp. 365–388). Mahwah, NJ: Lawrence Erlbaum Associates.

Fisher, C., & Tokura, H. (1996). Acoustic cues to grammatical structure in infant-directed speech: Cross-linguistic evidence. *Child Development*, *67*, 3192–3218.

Freudenthal, D., Pine, J. M., & Gobet, F. (2006). Modelling the development of children's use of optional infinitives in English and Dutch using MOSAIC. *Cognitive Science, 30*, 277–310.

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

Gerken, L. A. (1996). Prosody's role in language acquisition and adult parsing. *Journal of Psycholinguistic Research*, *25,* 345–356.

Gerken, L. A., Jusczyk, P. W., & Mandel, D. R. (1994). When prosody fails to cue syntactic structure: Nine-month-olds' sensitivity to phonological vs. syntactic phrases. *Cognition, 51*, 237–265.

Gleitman, L. & Wanner, E. (1982). Language acquisition: The state of the state of the art. In E. Wanner & L. Gleitman (Eds*.), Language acquisition: The state of the art* (pp. 3–48). Cambridge, UK: Cambridge University Press.

Gómez, R. L., & Gerken, L. A. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*, 109–135.

Gómez, R. L., & Gerken, L. A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*, 178–186.

Green, T. R. G. (1979). The necessity of syntax markers: Two experiments with artificial languages. *Journal of Verbal Learning and Verbal Behavior*, *18,* 481–496.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences*, *3*, 323–328.

Jusczyk, P. W., & Kemler-Nelson, D. G. (1996). Syntactic units, prosody, and psychological reality during infancy. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 389–408). Mahwah, NJ: Lawrence Erlbaum Associates.

Karmiloff-Smith, A. (1979). *A functional approach to child language: A study of determiners and reference*. Cambridge, UK: Cambridge University Press.

Kelly, M. H. (1988). Phonological biases in grammatical category shifts. *Journal of Memory and Language, 27*, 343–358.

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

Kemler-Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Wright Cassidy, K. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, *16*, 55–68.

Korman, M. (1984). Adaptive aspects of maternal vocalization in differing contexts at ten weeks. *First Language*, *5*, 44–45.

Kuhl, P. K. (1999). Speech, language, and the brain: Innate preparation for learning. In M. Konishi & M. Hauser (Eds.), *Neural mechanisms of communication* (pp. 419–450). Cambridge, MA: MIT Press.

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

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Maratsos, M., & Chalkley, M. A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K. Nelson (Ed.), *Children's language* (Vol. 2, pp. 127–214). New York: Gardner Press.

Maslen, R., Theakston, A., Lieven, E., & Tomasello, M. (2004) A dense corpus study of past tense and plural overregularization in English. *Journal of Speech, Language and Hearing Research, 47,* 1319–1333.

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.

McDonald, J. L., & Plauche, M. (1995). Single and correlated cues in an artificial language learning paradigm. *Language and Speech*, *38*, 223–236.

Mehler, J., Jusczyk, P. W., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*, 143–178.

Meier, R. P., & Bower, G. H. (1986). Semantic reference and phrasal grouping in the acquisition of a miniature phrase structure language. *Journal of Memory and Language*, *25*, 492–505.

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

Monaghan, P., & Christiansen, M. H. (2008). Integration of multiple probabilistic cues in syntax acquisition. In H. Behrens (Ed.), *Trends in corpus research: Finding structure in data* (pp. 139–163) (TILAR Series). Amsterdam: John Benjamins.

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological–distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology, 55*, 259–305.

Morgan, J. L. (1996). Prosody and the roots of parsing. *Language and Cognitive Processes*, *11*, 69–106.

Morgan, J. L., & Demuth, K. (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*, *19*, 498–550.

Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development, 66*, 911–936.

Morgan, J. L., Shi., R., & Allopenna, P. (1996). Perceptual bases of grammatical categories. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. (pp. 263–283). Mahwah, NJ: Lawrence Erlbaum Associates.

Newport, E. L., Gleitman, H., & Gleitman, L. R. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition* (pp. 109–149). Cambridge, UK: Cambridge University Press.

Onnis, L., & Christiansen, M. H. (2008). Lexical categories at the edge of the word. *Cognitive Science*, *32*, 184–221.

Pallier, C., Christophe, A., & Mehler, J. (1997). Language-specific listening. *Trends in Cognitive Sciences*, *1*, 129–132.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

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

Rubel, E. W. (1985). Auditory system development. In G. Gottlieb & N. A. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life* (pp. 53–89). Norwood, NJ: Ablex.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, *12*, 110–114.

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

Sereno, J. A., & Jongman, A. (1995). Acoustic correlates of grammatical class. *Language and Speech*, *38*, 57–76.

Shady, M., & Gerken, L. A. (1999). Grammatical and caregiver cues in early sentence comprehension. *Journal of Child Language*, *26*, 163–175.

Shafer, V. L., Shucard, D. W., Shucard, J. L., & Gerken, L. A. (1998). An electrophysiological study of infants' sensitivity to the sound patterns of English speech. *Journal of Speech, Language, and Hearing Research*, *41*, 874–886.

Shi, R., Morgan, J., & Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *Journal of Child Language*, *25*, 169–201.

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

Snedeker, J., & Trueswell, J. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology, 49(3)*, 238–299.

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, *2*, 191–196.

Steyvers, M., & Tenenbaum, J. (2005). The large scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science, 29*, 41–78.

Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences*, *4*, 156–163.

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and Brain Sciences*, *16*, 495–552.

Valian, V., & Coulson, S. (1988). Anchor points in language learning: The role of marker frequency. *Journal of Memory and Language*, *27*, 71–86.

Valian, V., & Levitt, A. (1996). Prosody and adults' learning of syntactic structure. *Journal of Memory and Language, 35*, 497–516.

Weissenborn, J., & Höhle, B. (Eds.) (2001). *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition.* Philadelphia, PA: John Benjamins.

Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology, 50*, 509–535.

Table 5.1 The Stochastic Phrase-Structure Grammar Used to Generate Training Corpora for Simulations 1–4

S → Imperative [0.1] | Interrogative [0.3] | Declarative [0.6]
Declarative → NP VP [0.7] | NP-ADJ [0.1] | That-NP [0.075] | You-P [0.125]
NP-ADJ → NP *is/are* adjective That-NP → *that/those is/are* NP You-P → *you are* NP
Imperative → VP
Interrogative → Wh-Question [0.65] | Aux-Question [0.35] Wh-Question → *where/who/what is/are* NP [0.5] | *Where/who/what do/does* NP VP [0.5] Aux-Question → *do/does* NP VP [0.33] | *Do/does* NP *wanna* VP [0.33] | *is/are* NP adjective [0.34]
NP → *a/the* N-sing/N-plur
VP → V-int | V-trans NP

Table 5.2 Phonological Cues that Distinguish between Lexical Categories

Nouns and Verbs
Nouns have more syllables than verbs (Kelly, 1992)
Bisyllabic nouns have 1st syllable stress, verbs tend to have 2nd syllable stress (Kelly & Bock, 1988)
Inflection -ed is pronounced /d/ for verbs, /@d/ or /Id/ for adjectives (Marchand, 1969)
Stressed syllables of nouns have more back vowels than front vowels. Verbs have more front vowels than back vowels (Sereno & Jongman, 1990)
Nouns have more low vowels, verbs have more high vowels (Sereno & Jongman, 1990)
Nouns are more likely to have nasal consonants (Kelly, 1992)
Nouns contain more phonemes per syllable than verbs (Kelly, 1996)

Table 5.2 Phonological Cues that Distinguish between Lexical Categories

Nouns and Verbs
Function and Content Words
Function words have fewer syllables than content words (Morgan, Shi & Allopenna, 1996)
Function words have minimal or null onsets (Morgan, Shi & Allopenna, 1996)
Function word onsets are more likely to be coronal (Morgan, Shi & Allopenna, 1996)
/D/ occurs word-initially only for function words (Morgan, Shi & Allopenna, 1996)
Function words have reduced vowels in the first syllable (Cutler, 1993)
Function words are often unstressed (Gleitman & Wanner, 1982)

Figure 5.1 The general architecture of the simple-recurrent network (SRN) employed across simulations. An input layer representing information relevant for individual words along with an utterance boundary marker feeds into a hidden layer, and then to an output that predicts information relevant to the following word in a corpus. The hidden layer copies itself to a context layer, which supplies a limited memory for past words.

Figure 5.2 Comparison of learning performance for different cue combinations in Simulation 1, showing that multiple-cue integration leads to (A) better learning (as measured by the lowest error obtained on the test corpus), (B) faster learning (measured in terms of the amount of training needed to surpass the performance of the trigram model), and (C) more uniform learning (as indicated by less variance across the performance of the different instances of the network). (Error bars = S.E.M.)

Figure 5.3 The effect of prosody and grammatical markers on human and SRN sentence processing. (A) Percent correct picture identification by 2-year-olds in the prosody condition of the Shady and Gerken (1999) experiment, with pauses inserted early, late, or in the unnatural position between the determiner and the noun. (B) Total activation of nouns by the SRN when exposed to the same prosodic manipulation as the human children. (C) Picture identification performance in the grammatical marker condition in Shady and Gerken (1999), involving a grammatical, nonsense, or ungrammatical word before the target noun. (D) Matching SRN activation of nouns for the same three types of grammatical markers. (Error bars = S.E.M.)

Figure 5.4 Speed of learning for networks trained with or without prenatal exposure to prosody and gross-level properties of phonology. (Error bars = S.E.M.)

Figure 5.5 Speed of learning for networks trained with or without distractor cues. (Error bars = S.E.M.)

Figure 5.6 Performance of the network models trained on full-blown child-directed speech. (A) Test performance for networks provided only with distributional cues and networks provided with both phonological and distributional cues. (B) Results of the discriminant analyses, comparing the ability of the two types of networks to place themselves in a “noun state” and a “verb state” when processing novel nouns and verbs, respectively. (Error bars = S.E.M.)