Title
Integrating Distributional, Prosodic and Phonological Information in a Connectionist Model of Language Acquisition

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 23(23)

ISSN
1069-7977

Authors
Christiansen, Morten H.
Dale, Rick A.

Publication Date
2001

Peer reviewed

Integrating Distributional, Prosodic and Phonological Information in a Connectionist Model of Language Acquisition

Morten H. Christiansen†‡ (morten@siu.edu)
Rick A.C. Dale† (racdale@siu.edu)
†Department of Psychology
‡Department of Linguistics
Carbondale, IL 62901 USA

Abstract

Children acquire the syntactic structure of their native language with remarkable speed and reliability. Recent work in developmental psycholinguistics suggests that children may bootstrap grammatical categories and basic syntactic structure by exploiting distributional, phonological, and prosodic cues. However, these cues are probabilistic, and are individually unreliable. In this paper, we present a series of simulations exploring the integration of multiple probabilistic cues in a connectionist model. The first simulation demonstrates that multiple-cue integration promotes significantly better, faster, and more uniform acquisition of syntax. In a second simulation, we show how this model can also accommodate recent data concerning the sensitivity of young children to prosody and grammatical function words. Our third simulation illuminates the potential contribution of prenatal language experience to the acquisition of syntax through multiple-cue integration. Finally, we demonstrate the robustness of the multiple-cue model in the face of potentially distracting cues, uncorrelated with grammatical structure.

Introduction

Before children can ride a bicycle or tie their shoes, they have learned a great deal about how words are combined to form complex sentences. This achievement is especially impressive because children acquire most of this syntactic knowledge with little or no direct instruction. Nevertheless, mastering natural language syntax may be among the most difficult learning tasks that children face. In adulthood, syntactic knowledge can be characterized by constraints governing the relationship between grammatical categories of words (such as noun and verb) in a sentence. But acquiring this knowledge presents the child with a "chicken-and-egg" problem: the syntactic constraints presuppose the grammatical categories in terms of which they are defined; and the validity of grammatical categories depends on how far they support 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. But the bootstrapping problem solved by young children seems vastly more challenging, both because the constraints governing natural language are so intricate, and because young children do not have the intellectual capacity or explicit instruction available to the academic student. 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 the input that are particularly informative for acquiring grammatical information. Specifically, we hypothesize that integrating multiple probabilistic cues (phonological, prosodic and distributional) by perceptually attuned general-purpose learning mechanisms may hold the key to 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 paper, we first review empirical evidence suggesting that infants may use a combination of distributional, phonological and prosodic cues to bootstrap into language. We then report a series of simulations, demonstrating the efficacy of multiple-cue integration within a connectionist framework. Simulation 1 shows how multiple-cue integration results in better, faster and more uniform learning. Simulation 2 establishes that the trained three-cue networks are able to mimic the effect of grammatical and prosodic manipulations in a sentence comprehension study with 2-year-olds (Shady & Gerken, 1999). Simulation 3 reveals how prenatal exposure to gross-level phonological and prosodic input facilitates postnatal learning within the multiple-cue integration framework. Finally, Simulation 4 demonstrates that adding additional distracting cues, irrelevant to the syntactic acquisition task, does not hinder learning.

Cues Available for Syntax Acquisition

Although some kind of *innate* knowledge may play a role in language acquisition, it cannot solve the bootstrapping problem. Even with built-in abstract knowledge about grammatical categories and syntactic rules (e.g., Pinker, 1984), the bootstrapping problem remains formidable: children must map the right sound strings onto the right grammatical categories while determining the specific syntactic relations between these categories in their native language. Moreover, the item-specific nature of early syntactic productions challenges the usefulness of hypothesized innate grammatical categories

(Tomasello, 2000).

Language-external information may substantially contribute to language acquisition. Correlations between environmental observations relating prior semantic categories (e.g., objects and actions) and grammatical categories (e.g., nouns and verbs) may furnish a "semantic bootstrapping" solution (Pinker, 1984). However, given that children acquire linguistic distinctions with 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. Another extra-linguistic factor is cultural learning where children may imitate the pairing of linguistic forms and their conventional communicative functions (Tomasello, 2000). Nonetheless, to break down the linguistic forms into relevant units, it appears that cultural learning must be coupled with language-internal learning. Moreover, because the nature of language-external and innate knowledge is difficult to assess, it is unclear how this knowledge could be quantified: There are no computational models of how such knowledge might be applied to learning basic grammatical structure.

Though perhaps not the only source of information involved in bootstrapping the child into language, the potential contribution of *language-internal* information is more readily quantified. Our test of the multiple-cue hypothesis therefore focuses on the degree to which language-internal information (phonological, prosodic and distributional) may contribute to solving the bootstrapping problem.

Phonological information—including stress, vowel quality, and duration—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). Phonological information may also help distinguish between nouns and verbs. For example, nouns tend to be longer than verbs in English—a difference that even 3-year-olds are sensitive to (Cassidy & Kelly, 1991). These and other phonological cues, such as differences in stress placement in multi-syllabic words, have also been found to exist cross-linguistically (see Kelly, 1992, for a review).

Prosodic information provides cues for word and phrasal/clausal segmentation and may help uncover syntactic structure (e.g., Morgan, 1996). Acoustic analyses suggest that differences in pause length, vowel duration, and pitch indicate phrase boundaries in both English and Japanese child-directed speech (Fisher & Tokura, 1996). Infants seem highly sensitive to such language-specific prosodic patterns (for reviews, see e.g., Jusczyk, 1997; Morgan, 1996)—a sensitivity that may start in utero (Mehler et al., 1988). Prosodic information also improves sentence comprehension in two-year-olds (Shady & Gerken, 1999). Results from an artificial language learning experiment with adults show that prosodic marking of syntactic phrase boundaries facilitates learning (Morgan, Meier & Newport, 1987). Unfortunately, prosody is partly affected by a number of non-syntactic factors, such as breathing patterns (Fernald & McRoberts, 1996), resulting in an imperfect mapping between prosody and syntax. Nonetheless, infants' sensitivity to prosody provides a rich potential source of syntactic information (Morgan, 1996).

None of these cues in isolation suffice to solve the bootstrapping problem; rather, they must be integrated to overcome the partial reliability of individual cues. Previous connectionist simulations by Christiansen, Allen and Seidenberg (1998) have pointed to efficient and robust learning methods for multiple-cue integration in speech segmentation. Integration of phonological (lexical stress), prosodic (utterance boundary), and distributional (phonetic segment sequences) information resulted in reliable segmentation, outperforming the use of individual cues. The efficacy of multiple-cue integration has also been confirmed in artificial language learning experiments (e.g., McDonald & Plauche, 1995).

By one year, children's perceptual attunement is likely to allow them to utilize language-internal probabilistic cues (for reviews, see e.g., Jusczyk, 1997; Kuhl, 1999). For example, infants 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, Shucard, Shucard & Gerken, 1998). Young infants can detect differences in syllable number among isolated words (Bijeljac, Bertoncini & Mehler, 1993)—a possible cue to noun/verb differences. Moreover, infants are accomplished distributional learners (e.g., Saffran, Aslin & Newport, 1996), and importantly, they are capable of multiple-cue integration (Mattys, Jusczyk, Luce & Morgan, 1999). When solving the bootstrapping problem children are also likely to benefit from specific properties of child-directed speech, such as the predominance of short sentences (Newport, Gleitman & Gleitman, 1977) and the cross-linguistically more robust prosody (Kuhl et al., 1997).

This review has indicated the range of language-internal cues 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).

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 paper are therefore intended as a first step toward a computational approach to multiple-cue integration, seeking to test the potential advantages of this approach to syntactic acquisition. Based on our previous experience with modeling multiple-cue integration in speech segmentation (Christiansen et al., 1998), we used a simple recurrent network (SRN; Elman, 1990) to model the integration of multiple cues. The networks were trained on corpora of artificial child-directed speech generated by a well-motivated grammar that includes three probabilistic cues to grammatical structure: word length,

lexical stress and pitch. 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.

Method

Networks Ten SRNs were used in each cue condition, with an initial weight randomization 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, and a constellation of cue units depending on its assigned cue condition. 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 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 45-50 across conditions. Each network had 80 hidden units and 80 context units.

Materials We constructed a complex grammar based on independent analyses of child-directed corpora (Bernstein-Ratner, 1984; Korman, 1984), and a study of child-directed speech by mother-daughter pairs (Fisher & Tokura, 1996). As illustrated in Table 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*).

Two basic cues were available to all networks. The fundamental distributional information inherent in the grammar could be exploited by all networks in this experiment. As a second basic cue, utterance-boundary pauses signalled 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 semi-reliable 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 were 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 signalled 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.

Table 1: **The Stochastic Phrase Structure Grammar Used to Generate Training Corpora**

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

Procedure Eight groups of networks, one for each combination of cues, were trained on corpora consisting of 10,000 sentences generated from the grammar. Each network within a group was trained on a different 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—and, importantly, this measure did not include predictions of cue information.

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

All networks achieved better performance than the standard bigram/trigram models ($p's < .0001$), suggesting that the networks had acquired knowledge of syntactic structure beyond the information associated with simple pairs or triples of words. The nets provided with phonological/prosodic cues achieved significantly better performance than base networks ($p's < .02$). Using trigram performance as criterion, all multiple-cue networks surpassed this level of performance faster than the base networks ($p's < .002$). Moreover, the three-cue networks were significantly faster than the single-cue networks ($p's < .001$). 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 base networks ($F(1,18) = 5.14, p < .04$).

Simulation 2: Sentence Comprehension in Two-Year-Olds

Simulation 1 provides evidence for the general feasibility of the multiple-cue integration approach. However, to further strengthen the model's credibility closer contact with relevant human data is needed. In the current simulation, we demonstrate that the three-cue networks from Simulation 1 are able to accommodate recent data showing that two-year-olds can integrate grammatical markers (function words) and prosodic cues in sentence comprehension (Shady & Gerken, 1999: Expt. 1). In this study, children heard sentences, such as (1), 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). Simulation 2 replicates this by using comparable stimuli, and assessing the noun activations.

Method

Networks Twelve networks from Simulation 1 were used in each prosodic condition. 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 represented by activating the utterance-boundary unit. Because these pauses probabilistically signal grammatically distinct utterances, the utterance-boundary unit provides a good 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 with 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

Shady and Gerken (1999) reported a significant effect of prosody on the picture selection task. The same was true for our networks ($F(2,33) = 1,253.07, p < .0001$). 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 did our networks ($F(2,70) = 69.85, p < .0001$), showing the most activation following the determiner, then for the nonsense word, and lastly for the ungrammatical word. This replication of the human data confers further support for Simulation 1 as a model of language acquisition by multiple-cue integration.

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 non-prenatal 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 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 non-prenatal 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 non-prenatal), 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-factor analysis of variance (ANOVA) with training condition (prenatal vs. non-prenatal) 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 benefitted significantly from training. An interaction between training conditions and degrees of training indicates that the prenatal networks learned significantly better than postnatal networks ($F(1,18) = 9.90, p < 0.01$). The exposure to prenatal input—void of any information about individual words—promotes better performance on the prediction task; thus providing computational support for the prenatal scaffolding hypothesis.

Simulation 4: Multiple-Cue Integration with Useful and Distracting Cues

A possible objection to the previous simulations is that our networks succeed at multiple-cue integration because they are "hand-fed" 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 the addition of the 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 intended to capture that some experience would likely 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 semi-reliable 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 non-prenatal). We repeated the two-factor ANOVA computed for Simulation 2, revealing a main effect for training condition ($F(1,18) = 4.76, p < 0.05$) 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 non-prenatal networks ($F(1,18) = 5.31, p < .05$).

To determine how the distractor cues may have 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 ($F(1,18) = 0.13, p = 0.72$). A further comparison between these non-prenatal 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 towards learning grammatical structure.

Conclusion

A growing bulk of evidence from developmental cognitive science has suggested that bootstrapping into language acquisition may be a process of integrating multiple sources of probabilistic information, each of which is individually unreliable, but jointly advantageous. However, what has so far been missing is a comprehensive demonstration of the computational feasibility of this approach. With the series of simulations reported here we have taken the first step toward establishing the computational advantages of multiple-cue integration. Simulation 1 demonstrated that providing SRNs with prosodic and phonological cues significantly improves their acquisition of syntactic structure—despite the fact that these cues are only partially reliable. The multiple-cue integration approach gains further support from Simulation 2, showing that the three-cue networks can mimic children's sensitivity to both prosodic and grammatical cues in sentence comprehension. The model also illustrates the potential value of prenatal exposure, since Simulation 3 revealed significant benefits for networks receiving such input. Finally, Simulation 4 provides evidence for the robustness of our neural network model, since highly unreliable cues did not interfere with the integration process. This implementation of our model still exhibited significant performance advantages over networks not receiving cues at all. Moreover, all the network models consistently performed better than the statistical benchmarks, the bigram and trigram models. This has important theoretical implications because it suggests that the SRNs acquired complex knowledge of grammatical structure and not merely simple two- or three-word co-occurrence statistics. Overall, the simulation results presented in this paper provide support not only for the multiple-cue integration approach in general, but also for a connectionist approach to the integration of distributional, prosodic and phonological information in language acquisition.

References

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.

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

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

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.

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.

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.

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

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

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

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., Stolyarova, E.I., Sundberg, U. & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, *277*, 684–686.

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.

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

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.

Newport, E.L., Gleitman, H. & Gleitman, L.R. (1977). Mother, Id 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.

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

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*. Norwood, NJ: Ablex.

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

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

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