Phonological and Distributional Cues in Syntax Acquisition: Scaling up the Connectionist Approach to Multiple-Cue Integration

Florencia Reali (fr34@cornell.edu)
Department of Psychology; Cornell University; Ithaca, NY 14853 USA
Morten H. Christiansen (mhc27@cornell.edu)
Department of Psychology; Cornell University; Ithaca, NY 14853 USA
Padraic Monaghan (Padraic.Monaghan@warwick.ac.uk)
Department of Psychology, University of Warwick; Coventry, CV4 7AL, UK

Abstract

Recent work in developmental psycholinguistics suggests that children may bootstrap grammatical categories and basic syntactic structure by exploiting distributional, phonological, and prosodic cues. Previous connectionist work has indicated that multiple-cue integration is computationally feasible for small artificial languages. In this paper, we present a series of simulations exploring the integration of distributional and phonological cues in a connectionist model trained on a full-blown corpus of child-directed speech. In the first simulation, we demonstrate that the connectionist model performs very well when trained on purely distributional information represented in terms of lexical categories. In the second simulation we demonstrate that networks trained on distributed vectors incorporating phonetic information about words also achieve a high level of performance. Finally, we employ discriminant analyses of hidden unit activations to show that the networks are able to integrate phonological and distributional cues in the service of developing highly reliable internal representations of lexical categories.

Introduction

Mastering natural language syntax may be one of the most difficult learning tasks that children face. This achievement is especially impressive given that children acquire most of this syntactic knowledge with little or no direct instruction. In adulthood, syntactic knowledge can be characterized by constraints governing the relationship between grammatical categories and words (such as noun and verb) in a sentence. The syntactic constraints presuppose the grammatical categories in terms of which they are defined; but the validity of grammatical categories depends on how far they support syntactic constraints. Thus, acquiring syntactic knowledge presents the child with a “bootstrapping” problem.

Bootstrapping into language seems to be vastly 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 adults. Yet, children solve this “chicken-and-egg” problem surprisingly well: Before they can tie their shoes they have learned a great deal about how words are combined to form complex sentences. Determining how children accomplish this challenging task remains an open question in cognitive science.

Some, perhaps most, aspects of syntactic knowledge have to be acquired from mere exposure. Acquiring the specific words and phonological structure of a language requires exposure to a significant corpus of language input. In this context, distributional cues constitute an important source of information for bootstrapping language structure (for a review, see Redington & Chater, 1998). By eight months, infants have access to powerful mechanisms to compute the statistical properties of the language input (Saffran, Aslin and Newport, 1996). By one year, children’s perceptual attunement is likely to allow them to use language-internal probabilistic cues (for reviews, see e.g. Jusczyk, 1997). For example, children appear to be sensitive to the acoustic differences reflected by the number of syllables in isolated words (Shi, Werker & Morgan, 1999), and the relationship between function words and prosody in speech (Shafer, Shucard, Shucard & Gerken, 1998). Children are not only sensitive to distributional information, but they also are capable of multiple-cue integration (Mattys, Jusczyk, Luce & Morgan, 1999).

The multiple-cue integration hypothesis (e.g., Christiansen & Dale, 2001; Gleitman & Wanner, 1982; Morgan, 1996) suggests that integrating multiple probabilistic cues may provide an essential scaffolding for syntactic learning by biasing children toward aspects of the input that are particularly informative for acquiring grammatical structure. In the present study we focus on the integration of distributional and phonological cues using a connectionist approach.

In the remainder of this paper, we first provide a review of the empirical evidence suggesting that infants may use several different phonological cues to bootstrap into language. We then present a series of simulations demonstrating the efficacy of distributional cues for the acquisition of syntactic structure. In previous research (Christiansen & Dale, 2001), we have shown the advantages of multiple-cue models for the acquisition of grammatical structure in artificial languages. In this paper we are seeking to scale up this model, by training it on a complex corpus of child-directed speech. In the first simulation we show that simple recurrent networks trained on lexical categories are able to predict grammatical structure from the corpus. In the second simulation, we show that a network trained with phonetic information about the words in the corpus performed better in bootstrapping syntactic structure than a control network trained on random inputs. Finally, we analyze the networks' internal representations for lexical categories, and find that the networks are capable of integrating both phonetic and distributional information in the service of developing reliable representations for nouns and verbs.

Phonological cues to lexical categories

There are several phonological cues that individuate lexical categories. Nouns and verbs are the largest such categories, and consequently have been the focus of many proposals for distinctions in terms of phonological cues. Distinctions have also been made between function words and content words. Table 1 summarizes a variety of phonological cues that have been proposed to distinguish between different syntactic categories.

Corpus-based studies of English have indicated that distinctions between lexical categories based on each of these cues considered independently are statistically significant (Kelly, 1992). Shi, Morgan and Allopenna (1998) assessed the reliability of several cues when used simultaneously in a discriminant analysis of function/content words from small corpora of child-directed speech. They used several cues at the word level (e.g., frequency), the syllable level (e.g., number of consonants in the syllable), and the acoustic level (e.g., vowel duration) and produced 83%-91% correct classification for each of the mothers in the study. Durieux and Gillis (2001) considered a number of phonological cues for distinguishing nouns and verbs and, with an instance-based learning system correctly classified approximately 67% of nouns and verbs from a random sample of 5,000 words from the CELEX database (Baayen, Pipenbrock & Gulikers, 1995).

Cassidy and Kelly (1991) report experimental data indicating that phonological cues are used in lexical categorization. Participants were required to listen to a nonword and make a sentence including the word. The nonword stimuli varied in terms of syllable length. They found that longer nonwords were more likely to be placed in noun contexts, whereas shorter nonwords were produced in verb contexts.

Monaghan, Chater and Christiansen (in press) have shown that sets of phonological cues, when considered integratively, can predict variance in response times on naming and lexical decision for nouns and verbs. Words that are more typical of the phonological characteristics of their lexical category have quicker response times than words that share few cues with their category.

We accumulated a set of 16 phonological cues, based on the list in Table 1. Some entries in the Table generated more than one cue. For example, we tested whether reduced vowels occurred in the first syllable of the word, as well as testing for the proportion of reduced vowels throughout the word. We tested each cue individually for its power to discriminate between nouns and verbs from the 1000 most frequent words in a child-directed speech database (CHILDES, MacWhinney, 2000). There were 402 nouns and 218 verbs in our analysis. We conducted discriminant analyses on each cue individually, and found that correct classification was only just above chance for each cue. The best performance was for syllable length, with correct classification of 41.0% of nouns and 74.8% of verbs (overall, with equally weighted groups, 57.4% correct classification). When the cues were considered jointly, 92.5% of nouns and 41.7% of verbs were correctly classified (equal-weighted-group accuracy was 74.7%). The cues, when considered together resulted in accurate classification for nouns, but many verbs were also incorrectly classified as nouns (see Monaghan, Chater & Christiansen, submitted).

The discriminant analysis indicates that phonological information is useful for lexical categorization, but not sufficient without integration with cues from other sources. In the following simulations, we first show how a connectionist model is capable of learning aspects of syntactic structure from the distributional information derived from a corpus of child-directed speech. The

Table 1: 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)
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)

subsequent simulation and hidden unit analyses then explore how networks may benefit from integrating the kind of phonetic cues described above with distributional information.

Simulation 1: Learning syntactic structure using SRNs

We trained simple recurrent networks (SRN; Elman, 1990) to learn the syntactic structure present in a child-directed speech corpus. Previous research has shown that SRNs are able to acquire both simple (Christiansen & Chater, 1999; Elman, 1990) and slightly more complex and psychologically motivated artificial languages (Christiansen & Dale, 2001). An important outstanding question is whether these artificial-language models can be scaled up to deal with the full complexity and the general disorderliness of speech directed at young children. Here, we therefore seek to determine whether the advantages of distributional learning in the small-scale simulations will carry over to the learning of a natural corpus. Our simulations demonstrate that these networks are indeed capable of acquiring important aspects of the syntactic structure of realistic corpora from distributional cues alone.

Method

Networks Ten SRNs were used 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 localist representation of the lexical category of the incoming word. With a total of 14 different lexical categories and a pause marking boundaries between utterances, the network had 15 input units. The network was trained to predict the lexical category of the next word, and thus the number of output units was 15. Each network had 30 hidden units and 30 context units.

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 4-5 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%). Each word in the original corpus was replaced by a vector encoding the lexical category to which it belonged. 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).

Procedure The ten SRNs were trained on the corpus described above. Training consisted of 10 passes through the training corpus. Performance was assessed based on the networks ability to predict the next set of lexical categories given the prior context.

Results and discussion

Each lexical category was represented by the activation of a single unit in the output layer. 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 occurrence of lexical categories within an utterance (Christiansen & Chater, 1999). Thus, with c_i denoting the lexical category of *ith* word in the sentence we have the following relation:

$$P(c_p \mid c_1, c_2, \ldots c_{p-1}) = \text{Freq}(c_1, c_2, \ldots, c_{p-1}, c_p) \,/\, \text{Freq}(c_1, c_2, \ldots, c_{p-1})$$

where the probability of getting some member of a lexical category is conditional on the previous *p-1* categories.

To provide a statistical benchmark with which to compare the network performance, we “trained” bigram and trigram models on the Bernstein-Ratner corpus. These finite-state models borrowed from computational linguistics, provide a simple prediction method based on strings of two (bigrams) or three (trigrams) consecutive words.

We compared the full conditional probabilities with the network output probabilities and also with bigram and trigram predictions. The mean cosine between the full conditional probabilities for the test set and the predictions of the finite-state models were 0.81 for trigrams and 0.79 for bigrams. We found that the mean cosine between the full conditional probabilities and network output probabilities were comparable to the finite-state models’ predictions (mean cosine: 0.86 for the training set and mean cosine: 0.79 for the test set). Network predictions for the training set were better than bigram predictions ($p's < .00005$) and trigram predictions ($p's < .00005$). For the test set we found that the trigram model performed better than the networks ($p's < .00005$), but there was no difference between the performance of the bigram model and that of the networks ($p's < .52$).

Despite the complexity of child-directed speech – including the many false starts and other ‘ungrammatical’ constructions – Simulation 1 demonstrates that the SRN model is able to acquire much of the syntactic structure in the corpus from a single cue: distributional information.

Although these results fit with results from computational linguistics where models are often trained on corpora encoded in the form of lexical categories, it is also clear that children are not provided directly with such “tagged” input. Rather, the child has to bootstrap both lexical categories and syntactic constraints concurrently. One way of doing this may be to combine distributional information with other kinds of cues. Therefore in the next simulation we trained the same type of network but with each word encoded not simply by fifteen lexical categories but instead by more than a thousand different vectors encompassing different types of phonological information.

Simulation 2: Phonological cues for syntactic acquisition

We here take a further step towards providing a more solid computational foundation for multiple-cue integration. In Simulation 1 we provided evidence of the efficacy of using SRNs for learning syntactic structure from the corpus. Our next aim is to determine the extent to which these networks are sensitive to the lexical category information present in a 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 mentioned above (Table 1). The second set of input was encoded using a random encoding. Possible performance differences in networks trained with these different input sets would be due to lexical category information available in the 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]. We used the same values for learning rate and momentum as in Simulation 1. Each input to the network contained a thermometer encoding for each of the 16 phonological cues. 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 for the different lexical categories similarly to Simulation 1. With the 14 different lexical categories and a pause marking boundaries between utterances, the networks had 15 output units. Each network furthermore was equipped with 88 hidden units and 88 context units.

Materials We used the same training and test corpora as in Simulation 1. Each word was encoded according to the following sixteen phonological cues: 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 utterance boundaries.

The random-input networks were trained using input for which we randomly permuted the multiple-cue vectors among all the words in the corpus. Thus, the random vector encoding a given word would be reassigned to an arbitrary 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 ten control networks were trained on the random vectors. The training consisted of a pass through the training corpus. We used the same ten random seeds for both simulation conditions. As in Simulation 1, 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 and discussion

As in Simulation 1, we recorded the networks’ output probabilities and computed the full conditional probability vectors for the two groups of networks. We compared the predictions of the phonetic-input networks with those of the random-input networks[1]. The mean cosine between the full conditional probabilities and the phonetic-input networks’ output probabilities was 0.75 for the training set and 0.69 for the test set. For the random-input networks, we found that the mean cosine was 0.73 for the training set, and 0.67

[1] Given the absence of explicit lexical category information in the input combined with the complexity and nature of the phonetic encoding in Simulation 2, network performance is not directly comparable with that of the bigram/trigram models. Thus, the seemingly better performance of the finite-state models (in terms of mean cosine) is somewhat deceptive.

for the test set. The phonetic-input networks were significantly better than the random-input networks at predicting the next combination of lexical categories, both for the training set (*p's* < .00005) and the test set (*p's* < .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.

Probing the internal representations

Simulation 2 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 2. The hidden unit activations were recorded for 200 novel nouns and 200 novel verbs occurring in unique sentences taken from other CHILDES corpora. 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 and Discussion

We first compared the two sets of networks. 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., submitted), 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[2].

Conclusions

A growing bulk of experimental evidence from developmental cognitive science has indicated that children are sensitive to and able to combine a host of different sources of information, and that this may help them overcome the syntactic bootstrapping problem. However, one important caveat regarding such multiple-cue integration is that the various sources of information are highly probabilistic, and each is unreliable when considered in isolation. Although some headway has been made in the investigation of possible computational mechanisms that may be able to integrate multiple probabilistic cues, this work has primarily been small in scale (e.g., Christiansen & Dale, 2001).

In this paper, we have presented two series of simulations aimed at taking the first steps towards scaling up connectionist models of multiple-cue integration to deal with the full complexity of natural speech directed at children. The results of Simulation 1 demonstrated that SRNs are capable of taking advantage of distributional information – despite the many grammatical inconsistencies found in child-directed speech. In Simulation 2, we

[2] It is possible to object that children do not only get cues that are relevant for syntactic acquisition. Christiansen and Dale (2001) specifically addressed this issue by providing networks with additional cues that did not correlate with syntax. They found that the networks were able to ignore such "distractor" cues and focus on the relevant cues.

expanded these results by showing that SRNs are also able to utilize the highly probabilistic information found in the 16 phonological cues in the service of syntactic acquisition. Our probing 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 results presented here underscore the computational feasibility of the multiple-cue integration hypothesis, in particular within a connectionist framework.

Acknowledgments

This research was supported by the Human Frontiers Science Program.

References

Baayen, R.H., Pipenbrock, R. & Gulikers, L. (1995). *The CELEX Lexical Database* (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

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

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

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

Christiansen, M.H. & Dale, R.A.C. (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.

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

Durieux, G. & Gillis, S. (2001). Predicting grammatical classes from phonological cues: An empirical test. In J. Weissenborn & B. Höhle (Eds.) *Approaches to Bootstrapping: Phonological, Lexical, Syntactic and Neurophysiological Aspects of Early Language Acquisition* Volume 1. Amsterdam: John Benjamins.

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

Gleitman, L.R. & Wanner, E. (1982). Language acquisition: The state of the state of the art. In E. Wanner & L.R. Gleitman (Eds.) *Language Acquisition: The State of the Art* (pp.3-48). Cambridge: Cambridge University Press.

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

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

Kelly, M.H. (1996). The role of phonology in grammatical category assignment. In J.L. Morgan & K. Demuth (Eds.) *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kelly, M.H. & Bock, J.K. (1988). Stress in time. *Journal of Experimental Psychology: Human Perception & Performance, 14*, 389-403.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk* (3rd Edition). Mahwah, NJ: Lawrence Erlbaum Associates.

Marchand, H. (1969). *The Categories and Types of Present-day English Word-formation* (2nd Edition). Munich, Germany: C.H. Beck'sche Verlagsbuchhandlung.

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.

Monaghan, P., Chater, N. & Christiansen, M.H. (in press). Inequality between the classes: Phonological and distributional typicality as predictors of lexical processing. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.

Monaghan, P., Chater, N. & Christiansen, M.H. (submitted). Differential Contributions of Phonological and Distributional Cues in Language Acquisition.

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

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*. Mahwah, NJ: Lawrence Erlbaum Associates.

Redington, M., & Chater, N. (1998). Connectionist and statistical approaches to language acquisition: A distributional perspective. *Language and Cognitive Processes*, *13*, 129-191.

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. (1990). Phonological and form class relations in the lexicon. *Journal of Psycholinguistic Research*, **19**, 387-404.

Shaffer, 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.

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