Structure Dependence in Language Acquisition: Uncovering the Statistical Richness of the Stimulus

Florencia Reali (fr34@cornell.edu) and Morten H. Christiansen (mhc27@cornell.edu)
Department of Psychology; Cornell University; Ithaca, NY 14853 USA

Abstract

The *poverty of stimulus argument* is one of the most controversial arguments in the study of language acquisition. Here we follow previous approaches challenging the assumption of impoverished primary linguistic data, focusing on the specific problem of auxiliary fronting in polar interrogatives. We develop a series of child-directed corpus analyses showing that there is indirect statistical information useful for correct auxiliary fronting in polar interrogatives, and that such information is sufficient for producing grammatical generalizations even in the absence of direct evidence. We further show that there are simple learning devices, such as neural networks, capable of exploiting such statistical cues, producing a bias to correct *aux*-questions when compared to their ungrammatical counterparts. The results suggest that the basic assumptions of the poverty of stimulus argument need to be reappraised.

Introduction

How do children learn aspects of their language for which there appears to be no evidence in the input? This question lies at the heart of the most enduring and controversial debates in cognitive science. Ever since Chomsky (1965), it has been argued that the information in the linguistic environment is too impoverished for a human learner to attain adult competence in language without the aide of innate linguistic knowledge. Although this *poverty of the stimulus argument* (Chomsky, 1980; Crain & Pietroski, 2001) has guided most research in linguistics, it has proved to be much more contentious within the broader context of cognitive science.

The poverty of stimulus argument rests on certain assumptions about the nature of the input to the child, the properties of computational learning mechanisms, and the learning abilities of young infants. A growing bulk of research in cognitive science has begun to call each of these three assumptions into question. Thus, whereas the traditional nativist perspective suggests that statistical information may be of little use for syntax acquisition (e.g., Chomsky, 1957), recent research indicates that distributional regularities may provide an important source of information for syntactic bootstrapping (e.g., Mintz, 2002; Redington, Chater and Finch, 1998)—especially when integrated with prosodic or phonological information (e.g., Christiansen & Dale, 2001; Morgan, Meier & Newport, 1987). And while the traditional approach only tends to consider learning in highly simplified forms, such as "move the first occurrence of *X* to *Y*", progress in statistical natural language processing and connectionist modeling has revealed much more complex learning abilities of potential relevance for language acquisition (e.g., Lewis & Elman, 2001). Finally, little attention has traditionally been paid to what young infants may be able to learn, and this may be problematic given that recent research has demonstrated that even before one year of age, infants are quite competent statistical learners (Saffran, Aslin & Newport, 1996—for reviews, see Gómez & Gerken, 2000; Saffran, 2003).

These research developments suggest the need for a reappraisal of the poverty of stimulus argument, centered on whether they together can answer the question of how a child may be able to learn aspects of linguistic structure for which innate knowledge was previously thought to be necessary. In this paper, we approach this question in the context of structure dependence in language acquisition, specifically in relation to auxiliary fronting in polar interrogatives. We first outline the poverty of stimulus debate as it has played out with respect to forming grammatical questions with auxiliary fronting. It has been argued that the input to the child does not provide enough information to differentiate between correct and incorrect auxiliary fronting in polar interrogatives (Chomsky in Piatelli-Palmarini, 1980). In contrast, we conduct a corpus analysis to show that there is sufficiently rich statistical information available in child-directed speech for generating correct *aux*-questions—even in the absence of any such constructions in the corpus. We additionally demonstrate how the same approach can be applied to explain results from studies of auxiliary fronting in 3- to 5-year-olds (Crain & Nakayama, 1987). Whereas, the corpus analyses indicate that there is rich statistical information available in the input, it does not show that there are learning mechanisms capable of utilizing such information. We therefore conduct a set of connectionist simulations to illustrate that neural networks are capable of using statistical information to distinguish between correct and incorrect *aux*-questions. In the conclusion, we discuss our results in the context of recent infant learning results.

The Poverty of Stimulus and Structure Dependence in Auxiliary Fronting.

Children only hear a finite number of sentences, yet they learn to speak and comprehend sentences drawn from a language that can contain an infinite number of sentences. The poverty of stimulus argument suggests that children do not have enough data during the early stages of their life to learn the syntactic structure of their language. Thus, learning a language involves the correct generalization of grammatical structure when insufficient data is available to children. The possible weakness of the argument lies in the difficulty to assess the input, and in the imprecise and intuitive definition of 'insufficient data'.

One of the most used examples to support the poverty of stimulus argument concerns auxiliary fronting in polar interrogatives. Declaratives are turned into questions by fronting the correct auxiliary. Thus, for example, in the declarative form '*The man who is hungry is ordering dinner'* it is correct to front the main clause auxiliary as in 1, but fronting the subordinate clause auxiliary produces an ungrammatical sentence as in 2 (Chomsky, 1965).

1. *Is the man who is hungry ordering dinner?*
2. **Is the man who hungry is ordering dinner?*

Children can generate two types of rules: a structure-independent rule where the first *'is'* is moved; or the correct structure-dependent rule, where only the movement of the *'is'* from the main clause is allowed. Crucially, children do not appear to go through a period when they erroneously move the first *is* to the front of the sentence (e.g., Crain & Nakayama, 1987). It has moreover been asserted that a person might go through much of his or her life without ever having been exposed to the relevant evidence for inferring correct auxiliary fronting (Chomsky, in Piatelli-Palmarini, 1980).

The purported absence of evidence in the primary linguistic input regarding auxiliary fronting in polar interrogatives is not without debate. Intuitively, as suggested by Lewis & Elman (2001), it is perhaps unlikely that a child would reach kindergarten without being exposed to sentences such as 3-5.

3. *Is the boy who was playing with you still there?*
4. *Will those who are hungry raise their hand?*
5. *Where is the little girl full of smiles?*

These examples have an auxiliary verb within the subject NP, and thus the auxiliary that appears initially would not be the first auxiliary in the declarative, providing evidence for correct auxiliary fronting. Pullum & Scholz (2002) explored the presence of auxiliary fronting in polar interrogatives in the Wall Street Journal (WSJ). They found that at least five crucial examples occur in the first 500 interrogatives. These results suggest that the assumption of complete absence of evidence for correct auxiliary fronting is overstated. Nevertheless, it has been argued that the WSJ corpus is not a good approximation of the grammatical constructions that young children encounter and thus it cannot be considered representative of the primary linguistic data. Indeed, studies of the CHILDES corpus show that even though interrogatives constitute a large percentage of the corpus, relevant examples of auxiliary fronting in polar interrogatives represent less than 1% of them (Legate & Yang, 2002).

Although the direct evidence for auxiliary fronting in polar interrogatives may be too weak to be helpful in acquisition—as suggested by Legate & Yang (2002)—other more indirect sources of statistical information may provide sufficient basis for making the appropriate grammatical generalizations. Recent connectionist simulations provide preliminary data in this regard. Lewis & Elman (2001) trained simple recurrent networks (SRN; Elman, 1990) on data from an artificial grammar that generated questions of the form 'AUX NP ADJ?' and sequences of the form 'A_i NP B_i' (where A_i and B_i represent a variety of different material) but no relevant examples of polar interrogatives. The SRNs were better at making predictions for correct auxiliary fronting compared to those with incorrect auxiliary fronting. This indicates that even without direct exposure to relevant examples, the statistical structure of the input nonetheless provides useful information applicable to auxiliary fronting in polar interrogatives.

However, the SRNs in the Lewis & Elman simulation studies were exposed to an artificial grammar without the complexity and noisiness that characterizes actual child-directed speech. The question thus remains whether the indirect statistical regularities in an actual corpus of child-directed speech are strong enough to support grammatical generalizations over incorrect ones—even in the absence of direct examples of auxiliary fronting in polar interrogatives in the input. Next, in our first experiment, we conduct a corpus analysis to demonstrate that the indirect statistical information available in a corpus of child-directed speech is indeed sufficient for making the appropriate grammatical generalizations in questions involving auxiliary fronting.

Experiment 1: Measuring Indirect Statistical Information Relevant for Auxiliary Fronting

Even if children only hear a few relevant examples of polar interrogatives, they may nevertheless be able to rely on indirect statistical cues for learning the correct structure. In order to assess this hypothesis, we trained bigram and trigram models on the Bernstein-Ratner (1984) corpus of child-directed speech and then tested the likelihood of novel example sentences. The test sentences consisted of correct polar interrogatives (e.g. *Is the man who is hungry ordering dinner?*) and incorrect ones (e.g. *Is the man who hungry is ordering dinner?*)—neither of which were present in the training corpus. We reasoned that if indirect statistical information provides a possible cue for generalizing correctly to the grammatical *aux*-questions, then we should find a difference in the likelihood of these two alternative hypotheses.

Bigram/trigram models are simple statistical models that use the previous one/two word(s) to predict the next one. Given a string of words or a sentence it is possible to compute the associated cross-entropy for that string of words according to the bigram/trigram model trained on a particular corpus (from Chen & Goodman, 1996). Thus, given two alternative sentences we can compare the probability of each of them as indicated by their associated cross-entropy as computed in the context of a particular corpus. Specifically, we can compare the two alternative generalizations of doing auxiliary fronting in polar interrogatives, comparing the cross-entropy associated with grammatical (e.g., *Is the man who is in the corner smoking?*) and ungrammatical forms (e.g., *Is the man who in the corner is smoking*). This will allow us to determine whether there may be sufficient indirect statistical information available in actual child-directed speech to decide between these two forms. Importantly, the Bernstein-Ratner corpus contains no examples of auxiliary fronting in polar interrogatives. Our hypothesis is therefore that the corpus nonetheless contains enough statistical information to decide between grammatical and ungrammatical forms.

Method

Models For the purpose of corpus analysis we used bigram and trigram models of language (see e.g., Jurafsky & Martin, 2000). The probability P(*s*) of a sentence was expressed as the product of the probabilities of the words (w_i) that compose the sentence, with each word probability conditional to the last *n*-1 words. Then, if $s = w_1 \ldots w_k$ we have:

$$P(s) = \Pi_i P(w_i|w^{i-1}{}_{i-n+1})$$

To estimate the probabilities of $P(w_i|w_{i-1})$ we used the *maximum likelihood* (ML) estimate for $P(w_i|w_{i-1})$ defined as (considering the bigram model):

$$P_{ML}(w_i|w_{i-1}) = P(w_{i-1}w_i) / P(w_{i-1}) = (c(w_{i-1}w_i)/Ns)/(c(w_{i-1})/Ns);$$

where Ns denote the total number of tokens and $c(\alpha)$ is the number of times the string α occurs in the corpus. Given that the corpus is quite small, we used the *interpolation smoothing technique* defined in Chen & Goodman (1996). The probability of a word (w_i) (or unigram model) is defined as:

$$P_{ML}(w_i) = c(w_i)/Ns;$$

The smoothing technique consists of the interpolation of the bigram model with the unigram model, and the trigram model with the bigram model. Thus, for the bigram model we have:

$$P_{interp}(w_i|w_{i-1}) = \lambda P_{ML}(w_i|w_{i-1}) + (1-\lambda)P_{ML}(w_i)$$

Accordingly for trigram models we have:

$$P_{interp}(w_i|w_{i-1}w_{i-2}) = \lambda P_{ML}(w_i|w_{i-1}w_{i-2}) + (1-\lambda)(\lambda P_{ML}(w_i|w_{i-1}) + (1-\lambda)P_{ML}(w_i)),$$

where λ is a value between 0 and 1 that determines the relative importance of each term in the equation. We used a standard $\lambda = 0.5$ so that all terms are equally weighted. We measure the likelihood of a given set of sentences using the measure of cross-entropy (Chen & Goodman, 1996). The cross-entropy of a set of sentences is defined as:

$$1/N_T \Sigma_i -\log_2 P(s_i) \quad \text{(where } s_i \text{ is the i}^{th} \text{ sentence).}$$

The cross-entropy value of a sentence is inversely correlated with the likelihood of it. Given a training corpus, and two sentences A and B we can compare the cross-entropy of both sentences and estimate which one is more probable according to the statistical information of the corpus. We used Perl programming in a Unix environment to implement the corpus analysis. This includes the simulation of bigram and trigram models and cross-entropy calculation and comparisons.

Materials We used the Bernstein-Ratner (1984) corpus of child-directed speech for our corpus analysis. It contains recorded speech from nine mothers speaking to their children over 4-5 months period when children were between the ages of 1 year and 1 month to 1 year and 9 months. This is a relatively small and very noisy corpus, mostly containing short sentences with simple grammatical structure. The following are some example sentences: *Oh you need some space; Where is my apple?; Oh. That's it*'.

Procedure We used the Bernstein-Ratner child-directed speech corpus as the training corpus for the bigram/trigram models. The models were trained on 10,082 sentences from the corpus (34,010 word tokens; 1,740 word types). We wanted to compare the cross-entropy of grammatical and ungrammatical polar interrogatives. For that purpose, we created two novel sets of sentences. The first one contained grammatically correct polar interrogatives and the second one contained the ungrammatical version of each sentence in the first set. The sentences were created using a random algorithm that selected words from the corpus, and created sentences according to syntactic and semantic constraints. We tried to prevent any possible bias in creating the test sentences. The test sets only contained relevant examples of polar interrogatives of the form: "*Is /* NP/ (who/that)/ *is* / A_i/ B_i?", where A_i and B_i represent a variety of different material including VP, PARTICIPLE,NP, PP, ADJP (e.g.: *"Is the lady who is there eating?*"; "*Is the dog that is on the chair black?*"). Each test set contained 100 sentences. We estimated the mean cross-entropy per sentence by calculating the average cross-entropy of the 100 sentences in each set. Then we compared the likelihood of pairs of grammatical and ungrammatical sentences by comparing their cross-entropy and choosing the version with the lower value. We studied the statistical significance of the results using paired t-test analyses.

Results

We found that the mean cross-entropy of grammatical sentences was lower than mean cross entropy of ungrammatical sentences. We performed a statistical analysis of the cross-entropy difference, considering all pairs of grammatical and ungrammatical sentences. The cross-entropy difference was highly significant ($t(99)$, $p<0.0001$) (see Table 1). These results show that grammatical sentences have a higher probability than ungrammatical ones. In order to compare each grammatical-ungrammatical pair of sentences, we defined the following criterion: When deciding between each grammatical vs. ungrammatical polar interrogative example, choose the one that has lower cross-entropy (the most probable one).

Table 1: Comparison of mean cross-entropy in Exp.1.

	Mean cross-entropy		Mean difference	t(99) p-value
	Gram.	Ungramm.		
Bigram	22.92	23.73	0.83	< 0.0001
Trigram	21.81	23.07	1.26	< 0.0001

A sentence is defined as correctly classified if the chosen form is grammatical. Using that criterion, we found that the percentage of correctly classified sentences using the bigram model is 92% and using the trigram model is 95%. Figure 1 shows the performance of the models according to the defined classification criterion. Of the 100 test sentences, the trigram model only misclassified the following five: *Is the lady who is here drinking?; Is the alligator that is standing there red?; Is the jacket that is on the chair lovely?; Is the one that is in the kitchen scared?; Is the phone that is in the office purple?*

The bigram model in addition to the above five sentences also misclassified the next three test sentences: *Is the bunny that is in the car little; Is the baby who is in the castle eating?; Is the bunny that is sleeping black?*

It is possible to calculate the probability of a sentence from the cross-entropy value. Figure 2 shows the comparison of mean probability of grammatical and ungrammatical sentences. We found that the mean probability of grammatical polar interrogatives is almost twice the mean probability of ungrammatical polar interrogatives according to the bigram model and it is more than twice according to the trigram model.

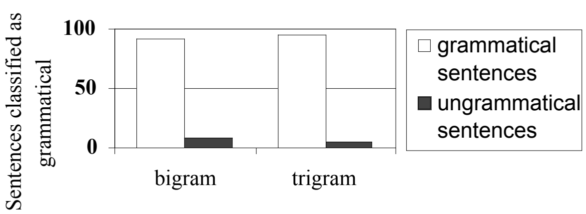

Figure 1: Number of sentences classified correctly (white bars) and incorrectly as grammatical (gray bars)

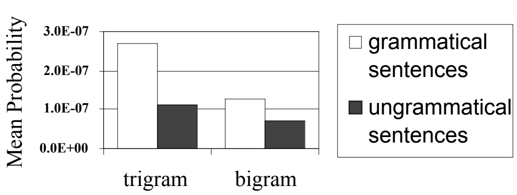

Figure 2: Mean probability of grammatical sentences vs. mean probability of ungrammatical sentences.

Experiment 2: Testing Sentences with Auxiliary Fronting Produced by Children

Although Experiment 1 shows that there is sufficient indirect statistical information available in child-directed speech to differentiate reliably between the grammatical and ungrammatical *aux*-questions that we had generated, it could be argued that the real test for our approach is whether it works for actual sentences produced by children. We therefore tested our models on a small set of sentences elicited from children under experimental conditions.

Crain & Nakayama (1987) conducted an experiment designed to elicit complex *aux*-questions from 3- to 5-year-old children. The children were involved in a game in which they asked questions to Jabba the Hutt, a creature from Star Wars. During the task the experimenter gives an instruction to the child: *'Ask Jabba if the boy who is watching Mickey Mouse is happy'*. Children produced sentences like a) *'Is the boy who is watching Mickey Mouse happy?'* but they never produced sentences like b) *'Is the boy who watching Mickey Mouse is happy?'*. The authors concluded that the lack of structure-independent errors suggested that children entertain only structure-dependent hypotheses, supporting the existence of innate grammatical structure.

Method

Models Same as in Experiment 1.

Materials Six example pairs were derived from the declarative sentences used in Crain & Nakayama[1](1987):

6. *The ball that the girl is sitting on is big*
7. *The boy who is unhappy is watching Mickey Mouse*
8. *The boy who is watching Mickey Mouse is happy*
9. *The boy who is being kissed by his mother is happy*
10. *The boy who was holding the plate is crying*
11. *The dog that is sleeping is on the blue bench*

The grammatical and ungrammatical *aux*-questions were derived from the declaratives in 6-11. Thus, the sentence '*Is the dog that is sleeping on the blue bench?*' belonged to the grammatical test set whereas the sentence '*Is the dog that sleeping is on the blue bench?*' belonged to the ungrammatical test set. Consequently, grammatical and ungrammatical test sets contained 6 sentences each.

Procedure The bigram/trigram models were trained on the Bernstein-Ratner (1984) corpus as in Experiment 1, and tested on the material derived from Crain & Nakayama (1987).

Results

Consistently with Experiment 1, we found that the mean cross-entropy of grammatical sentences was significantly lower than the mean cross entropy of ungrammatical sentences both for bigram and trigram models (t(5) $p<0.013$ and $p<0.034$ respectively). Table 2 summarizes these results.

[1] As some of the words in the examples were not present in the Bernstein-Ratner corpus, we substitute them for semantically related ones: Thus, the words: "mother", "plate", "watching", "unhappy" and "bench" were replaced respectively by "mommy", "ball", "looking at", "crying" and "chair".

Using the classification criterion defined in Experiment 1, we found that all six sentences were correctly classified using the bigram model. That is, according to the distributional information of the corpus, all grammatical *aux*-questions were more probable than the ungrammatical version of them. When using the trigram model, we found that five out of six sentences were correctly classified.

Table 2: Comparison of mean cross-entropy in Exp.2.

	Mean cross-entropy		Mean difference	t(5) p-value
	Gram.	Ungramm.		
Bigram	26.99	27.89	0.90	< 0.013
Trigram	25.97	26.86	0.89	< 0.034

Experiment 3: Learning to Produce Correct Sentences with Auxiliary Fronting

While Experiments 1 and 2 establish that there is sufficient indirect statistical information in the input to the child to differentiate between grammatical and ungrammatical questions involving auxiliary fronting—including questions produced by children—it is not clear whether a simple learning device may be able to exploit such information to develop an appropriate bias toward the grammatical forms. To investigate this question, we took a previously developed SRN model of language acquisition (Reali, Christiansen & Monaghan, 2003), which had also been trained on the same corpus, and tested its ability to deal with *aux*-questions.

Previous simulations by Lewis & Elman (2001) have shown that SRNs trained on data from an artificial grammar were better at predicting the correct auxiliary fronting in *aux*-questions. An important question is whether the results shown using artificial-language models are still obtained when dealing with the full complexity and the general disorderliness of speech directed at young children. Thus, we seek to determine whether a previously developed connectionist model, trained on the same corpus, is sensitive to the same indirect statistical information that we have found to be useful in bigram/trigram models. SRNs are simple learning devices that have been shown to be sensitive to bigram/trigram information.

Method

Networks We used the same ten SRNs that Reali, Christiansen & Monaghan (2003) had trained to predict the next lexical category given the current one. These networks had 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. All networks were simulated using the Lens simulator in a Unix environment. No changes were made to the original networks and their parameters.

Materials We trained and tested the networks on the Bernstein-Ratner corpus similarly to the bigram/trigram models. Each word in the corpus corresponded to one of the 14 following lexical categories from CELEX database (Baayen, Pipenbrock & Gulikers, 1995): nouns, verbs, adjectives, numerals, infinitive markers, adverbs, articles, pronouns, prepositions, conjunctions, interjections, complex contractions, abbreviations, and proper names. Each word in the corpus was replaced by a vector encoding the lexical category to which it belonged. We used the two sets of test sentences used in Experiment 1, containing grammatical and ungrammatical polar interrogatives respectively. However, as the network was trained to predict lexical classes, some test sentences defined in Experiment 1 mapped onto the same string of lexical classes. For simplicity, we only considered unique strings, resulting in 30 sentences in each test set (grammatical and ungrammatical).

Procedure The ten SRNs from Reali, Christiansen & Monaghan (2003) were trained on one pass through the Bernstein-Ratner corpus. These networks were then tested on the *aux*-questions described above. To compare network predictions for the ungrammatical vs. the grammatical *aux*-questions, we measured the networks' mean squared error recorded during the presentation of each test sentence pair.

Results

We found that in all ten simulations the grammatical set of aux-questions produced a lower error compared to the ungrammatical ones. The mean squared-error per next lexical class prediction was 0.80 for the grammatical set and 0.83 in the ungrammatical one, this difference being highly significant ($t(29)$ $p < 0.005$). Out of the 30 test sentences, 27 grammatical sentences produced a lower error than its ungrammatical counterpart. On the assumption that sentences with the lower error will be preferred, SRNs would pick the grammatical sentences in 27 out of 30 cases.

It is worth highlighting that the grammatical and ungrammatical sets of sentences were almost identical, only differing on the position of the fronted "is" as described in Experiment 1. Thus, the difference in mean squared error is uniquely due to the words' position in the sentence. Despite the complexity of child-directed speech, these results suggest that simple learning devices such as SRNs are able to pick up on the existing distributional properties showed in Experiment 1. Moreover, differently to Experiment 1, here we explored the distributional information of the lexical classes alone and thus the network was blind to the possible information present in word-word co-occurrences.

Conclusion

In the corpus analyses, we showed that there is sufficiently rich statistical information available *indirectly* in child-directed speech for generating correct complex *aux*-questions—even in the absence of any such constructions in the corpus. We additionally demonstrated how the same approach can be applied to explain results from child-acquisition studies (Crain & Nakayama, 1987). These results indicate that indirect statistical information provides a possible cue for generalizing correctly to grammatical auxiliary fronting.

Whereas the corpus analyses indicate that there are statistical cues available in the input, it does not show that there are learning mechanisms capable of utilizing such information. However, previous results suggest that children are sensitive to the same kind of statistical evidence that we found in the present study. Saffran, Aslin & Newport (1996) demonstrated that 8 month-old children are particularly sensitive to transitional probabilities (similar to our bigram model). Sensitivity to transitional probabilities seems to be present across modalities, for instance in the segmentation of streams of tones (Saffran, Johnson, Aslin, & Newport, 1999). These and other results on infant statistical learning (see Gómez & Gerken, 2000) suggest that children have mechanisms for relying on implicit statistical information. SRNs are simple learning devices whose learning properties have been shown to be consistent with humans' learning abilities. Even though it was originally developed in a different context (Reali, Christiansen & Monaghan, 2003), our SRN model proved to be sensitive to the indirect statistical evidence present in the corpus, developing an appropriate bias toward the correct forms of *aux*-questions.

In conclusion, this study indicates that the poverty of stimulus argument may not apply to the classic case of auxiliary fronting in polar interrogatives, previously a corner stone in the argument for the innateness of grammar. Our results further suggest that the general assumptions of the poverty of stimulus argument may need to be reappraised in the light of the statistical richness of the language input to children.

Acknowledgments

This research was supported in part by a Human Frontiers Science Program Grant (RGP0177/2001-B).

References

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

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

Chen S.F. & Goodman J. (1996). An Empirical Study of Smoothing Techniques for Language Modeling. *Proceedings of the 34th Annual Meeting of ACL.*

Chomsky N. (1957). *Syntactic Structures.* Mouton and co.: The Hague.

Chomsky N. (1965). *Aspects of the Theory of Syntax.* Boston, MA: MIT Press.

Chomsky N. (1980). *Rules & Representation.* Cambridge, MA: MIT Press.

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.

Crain, S. & Nakayama, M. (1987). Structure dependence in grammar formation. *Language*. 63: 522-543.

Crain, S. & Pietroski, P. (2001). Nature, nurture and Universal Grammar. *Linguistics and Philosophy.* 24: 139-186.

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

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

Jurafsky, D. and Martin, J.H. (2000) *Speech and Language Processing.* Upper Saddle River, NJ: Prentice Hall.

Legate, J.A. & Yang, C. (2002) Empirical re-assessment of stimulus poverty arguments. *Linguistic Review.* 19: 151-162.

Lewis, J.D. & Elman, J.L. (2001). Learnability and the statistical structure of language: Poverty of stimulus arguments revisited. In *Proceedings of the 26th Annual Boston University Conference on Language Development* (pp. 359-370). Somerville, MA: Cascadilla Press.

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

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.

Piatelli-Palmarini, M. (ed.), 1980. *Language and Learning: The Debate between Jean Piaget and Noam Chomsky*. Cambridge, MA: Harvard University Press.

Pullum G.K. & Scholz B. (2002) Empirical assessment of stimulus poverty arguments. *Linguistic Review.* 19: 9-50.

Reali, F., Christiansen, M.H. & Monaghan, P. (2003). Phonological and Distributional Cues in Syntax Acquisition: Scaling up the Connectionist Approach to Multiple-Cue In *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 970-975). Mahwah, NJ: Lawrence Erlbaum.

Redington, M., Chater, N. & Finch, S. (1998). Distributional Information: A Powerful Cue for Acquiring Syntactic Categories. *Cognitive Science.* 22: 425-469.

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

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

Saffran, J.R., Johnson, E.K., Aslin, R. & Newport, E.L. (1999) Statistical learning of tone sequences by human infants and adults. *Cognition.* 70: 27-52.