Non-Linguistic Constraints on the Acquisition of Phrase Structure

Jenny R. Saffran (jsaffran@facstaff.wisc.edu)
Department of Psychology; 1202 W. Johnson Street
Madison, WI 53706 USA

Abstract

To what extent is linguistic structure learnable from statistical information in the input? One set of cues which might assist in the discovery of hierarchical phrase structure given serially presented input are the dependencies, or predictive relationships, present within phrases. In order to determine whether adult learners can use this statistical information, subjects were exposed to artificial languages which either contained or violated the kinds of dependencies which characterize natural languages. The results suggest that adults possess learning mechanisms which detect and utilize statistical cues to phrase and hierarchical structure. A second experiment contrasted the acquisition of these linguistic systems with the same grammars implemented as non-linguistic input (sequences of non-linguistic sounds or shapes). These findings suggest that constraints on the mechanisms which highlight the statistical cues which are most characteristic of human languages are not specifically tailored for language learning.

Introduction

While the idea that surface distributional patterns point to pertinent linguistic structures holds a distinguished place in linguistic history (e.g., Bloomfield, 1933; Harris, 1951), statistical learning has only recently re-emerged as a potential contributing force in language acquisition (though see Maratsos & Chalkley, 1980). This renewed interest in statistical learning has been fueled by developments in computational modeling, by the widespread availability of large corpora of child-directed speech, and most recently by empirical research demonstrating that human subjects can perform statistical language learning tasks in laboratory experiments. For example, computational algorithms can use the co-occurrence environments of words to discover form classes in large corpora (e.g., Cartwright & Brent, 1997; Finch & Chater, 1994; Mintz, 1996; Mintz, Newport, & Bever, 1995). Similarly, individual verb argument structures can be induced by models which tracks the co-occurrences of verbs and their arguments in the input (e.g., Schütze, 1994; Seidenberg & MacDonald, 1999). Extensive modeling work has also examined the statistical cues available for the discovery of word boundaries in continuous speech (e.g., Aslin, Woodward, LaMendola, & Bever, 1996; Brent & Cartwright, 1996; Cairns, Shillcock, Chater, & Levy, 1997; Christiansen, Allen, & Seidenberg, 1998; Perruchet & Vintner, 1998).

These models provide invaluable explorations of the extent to which statistical information is available, in princi-ple, to language learners equipped with the right distributional tools. But are humans such learners? A wealth of statistical cues are useless unless humans can detect and use them. In fact, recent research suggests that humans are extremely good at some statistical language learning tasks, such as word segmentation (e.g., Aslin, Saffran, & Newport, 1998; Goodsitt, Morgan & Kuhl, 1993; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996)

These results suggest that humans possess powerful statistical language learning mechanisms, which are likely to provide important contributions to the language learning process. At the same time, it is important to recognize that these mechanisms would not be useful in language acquisition unless they are somehow <u>constrained or biased</u> to perform only certain kinds of computations over certain kinds of input. The pertinent generalizations to be drawn from a linguistic corpus are awash in irrelevant information. Any learning device without the right architectural, representational, or computational constraints risks being sidetracked by the massive number of misleading generalizations available in the input (e.g., Gleitman & Wanner, 1982; Pinker, 1984). There are an infinite number of linguistically irrelevant statistics that an overly powerful statistical learner could compute: for example, which words are presented third in sentences, or which words follow words whose second syllable begins with *th* (e.g., Pinker, 1989).

One way to avoid this combinatorial explosion would be to impose constraints on statistical learning which perform only a subset of the logically possible computations. It is clear that learning in biological systems is limited by internal factors; there are species differences in which specific types of stimuli serve as privileged input (e.g., Garcia & Koelling, 1966; Marler, 1991). External factors also strongly bias learning, because input from structured domains consists of non-random information. In order for statistical learning accounts to succeed, learners must be similarly constrained: humans must be just the type of statistical learners who are best suited to acquire the type of input exemplified by natural languages, focusing on linguistically relevant statistics while ignoring the wealth of available irrelevant computations. Such constraints might arise from various sources, either specific to language or from more general cognitive and/or perceptual constraints on human learning.

We have recently begun to explore the possibility that statistical learning itself is constrained. This line of research focuses the acquisition of hierarchical phrase structure. While words are spoken and perceived serially, our representations of sequences of words are highly structured. Consider the sentence *The professor graded the exam.* This sequence of words cannot be grouped as follows – (*The*) (*professor graded the*) (*exam*) – because words that are part of the same phrase are separated. For example, determiners like *the* require nouns; separating these two types of words violates the dependency relations which are part of native speakers' knowledge of English. The correct groupings, (*The professor*) (*graded* (*the exam*)), reflect English phrase structure, which generates a non-linear hierarchically organized structure. Hierarchical phrase structure represents a fascinating learning problem, because the child must somehow arrive at non-linear structure which is richer than is immediately suggested by the serial structure of the input. How do children make this leap? Innate knowledge is one possibility; prosodic regularities may also serve to chunk the input into phrasal units (e.g., Morgan, Meier, & Newport, 1987).

Another type of potentially useful information in the input suggests a statistical learning solution (see also Morgan & Newport, 1981). Linguistic phrases contain dependency relations: the presence of some word categories depends on others. For example, English nouns can occur without determiners like *the* or *a*. However, if a determiner is present, a noun almost always occurs somewhere downstream. This type of predictive relationship, which characterizes basic phrase types, may offer a statistical cue that highlights phrasal units for learners. Research using artificial languages with phrase structure grammars suggests that adult and child learners can exploit predictive dependencies to discover phrases (Saffran, 2000).

These studies suggest that people are skilled statistical learners. But what about the constraints required for the successful acquisition of languages? A particularly useful type of constraint would bias statistical learning mechanisms to preferentially acquire the types of structures observed in natural languages. To address this issue, Experiment 1 assessed the extent to which adults' ability to acquire an artificial grammar is affected by the availability of predictive dependencies as cues to linguistic phrase structure.

Experiment 1

Participants. 40 monolingual English speaking undergraduates at the University of Rochester participated in this study, and were each paid $6. Subjects were randomly assigned to the two experimental conditions.

Materials. The artificial grammars were adapted from the language used by Morgan & Newport (1981). One of the languages used in this study was a small phrase structure grammar (Language P, for predictive), in which dependencies between word categories afforded predictive cues to phrases, as in natural languages (e.g., if D is present, A must be present). Importantly, attempts to impose English predictive structure onto the input would mislead learners, as the phrase structure of Language P was head-final while English is head-initial. The second language was equally complex in terms of its size and formal characteristics, but contained a phrase structure unlike natural languages (Language N, for non-predictive). This language did not contain predictive dependencies marking phrases. Rather, it was characterized by overarching optionality: the presence of one word type never predicted the presence of another, which generates statistical properties unlike natural languages (note, however, that this language still possesses phrase structure of a sort – the absence of one word type predicts the presence of another; e.g., if A is <u>not</u> present, D must be present). Each form class (A, C, etc.) included 2 - 4 nonsense words (e.g., the words for the A category were BIFF, RUD, HEP, and MIB).

Table 1. Phrase structure grammars for Experiments 1 - 2. Letters refer to word classes; items in parentheses are optional. In Language N, one member of each phrase type must be present; if both are present, they must be in the order described by the grammar.

Language P	**Language N**
$S \rightarrow AP + BP + (CP)$	$S \rightarrow AP + BP$
$AP \rightarrow A + (D)$	$AP \rightarrow (A) + (D)$
$BP \rightarrow CP + F$	$BP \rightarrow CP + F$
$CP \rightarrow C + (G)$	$CP \rightarrow (C) + (G)$

The language generated by Language N is no larger than the language generated by Language P. In fact, Language N contained fewer sentence types (nine) than Language P (twelve). Language N also had shorter sentences on average, presumably making it less daunting to the learner: Language P generated 60% more five word sentences than Language N, and only 40% as many three word sentences. For both languages, only sentence types with five or fewer words were used (eight types for Language P, nine for Language N). Both languages contained the same number of grammatical categories and vocabulary items.

Because the languages were so similar in terms of their non-structural attributes, comparison of learning outcomes is valid. Language P is larger, and contains longer sentences, which could make it more difficult to acquire. However, if predictiveness affects learning, then the structure of Language N might have hindered its acquisition. A trained speaker recorded a corpus of 50 sentences from each language, with uniformly descending prosody but no grouping cues to phrase structure. Subjects were randomly assigned to hear either Language P or Language N sentences. Following approximately 30 min. of auditory exposure to one of the

two languages (the corpus was repeated eight times during exposure), all participants received the same forced-choice test consisting of novel grammatical and ungrammatical sentences, in order to assess acquisition of the rules of the two languages. Importantly, attempts to impose English syntax on either language would hinder performance. No cues other than the statistical information mirroring the underlying phrase structure of the language were available to learners.

Results. Each group's overall performance was significantly better than would be expected by chance: for Language P, the total score was 22.8 out of a possible 30: $t(19) = 10.46$, $p < .0001$; for Language N, the total score was 20.55: $t(19) = 6.62$, $p < .0001$ (see Figure 1). The principal hypothesis of interest concerns differences in learning as a function of structural differences between the two languages. To address this question, the scores for the two language groups for items testing each of the five rules were submitted to an ANOVA. The main effect of Language (P versus N) was significant: $F(1, 38) = 4.2$, $p < .05$.

These findings suggest that humans may be constrained to learn most readily via exactly the types of cues present in languages. To the extent that this is the case, the structure of natural languages may have been shaped by the nature of human learning (e.g., Bever, 1970; Christiansen, 1994; Christiansen & Devlin, 1997; Morgan, Meier, & Newport, 1987; Newport, 1990). According to the constrained statistical learning hypothesis, the mechanisms underlying language acquisition are biased to assist learners in detecting the 'right' statistical properties of the input. On this view, human languages have been sculpted by human learning and processing mechanisms – thereby creating input which contains the types of properties most useful for human learners, and rendering a close match between constraints on human learning and constraints on natural language structure.

If learners are biased to preferentially acquire structures where one item predicts another, is this constraint on learning particularly tailored for linguistic input? Biases in learning mechanisms may develop tightly coupled with the particular structure they are designed to acquire. Alternatively, constraints to use predictive statistics may be more generally applied to other types of sequentially presented information, as suggested by the constrained statistical learning hypothesis. Constraints on statistical learning which are not specific to language acquisition, but rather on the acquisition and processing of serial information, may have shaped the structure of natural languages. Experiment 2 thus utilized non-linguistic stimuli from two different modalities: visual shapes and complex sounds. An additional condition included visual linguistic stimuli (written words). As in Experiment 1, we contrasted the acquisition of Language P and N.

Experiment 2

Participants. 154 monolingual English speaking undergraduates at the University of Wisconsin - Madison participated in this study participated in this study for course extra credit. Forty-four subjects were randomly assigned to the non-linguistic auditory condition, forty subjects to the non-linguistic visual condition, and thirty subjects to the linguistic visual condition. Within each exposure condition, half of the subjects were assigned to Language P and half were assigned to Language N.

Method. For the non-linguistic visual condition, we translated the Language P and N grammars shown above into languages of shapes (for a similar methodology, see Goldowsky, 1995). For example, consider the phrase structure rule: AP → A + (D). In the linguistic version of this language, the category A consisted of 4 nonsense words. In the visual version, the category A consisted of 4 distinct shapes (such as a red circle with stripes). Category membership could not be induced by shape similarity, unlike prior studies by Morgan & Newport (1981). Participants observed the language on a computer monitor: each shape was presented in the middle of the screen, one at a time, with the same timing parameters as the auditory linguistic stimuli used in Experiment 1. Following exposure, participants were tested using a forced-choice test analogous to the linguistic task, in which they saw two shape sequences, one after the other, and decided which shape sequence more closely approximated the exposure stimuli. The linguistic visual condition was identical to the non-linguistic visual condition except that the nonsense words from Experiment 1 were shown typed on the computer screen. In the non-linguistic auditory condition, we translated Language P and N into non-linguistic sounds drawn from the digitized bank of alert sounds provided with Windows 98. Each word corresponded to a different sound, chosen to be maximally discriminable (an ascending buzz, a chord, chimes, etc.). Sound "sentences" generated by Language P and N were presented auditorily at the same rate as the linguistic and visual stimuli. Following exposure, participants received the same forced choice test, translated into non-linguistic sounds. Neither of the two non-linguistic conditions contained any linguistic information.

Results. Each group's overall performance was significantly better than would be expected by chance: for Language P Non-linguistic auditory, Nonlinguistic visual, and Linguistic visual, $p < .0001$; for Language N Nonlinguistic visual, $p < .001$; for Language N Nonlinguistic auditory, $p < .001$; and for Language N Linguistic visual, $p < .05$ (see Figure 1). As in Experiment 1, the principal hypothesis concerns
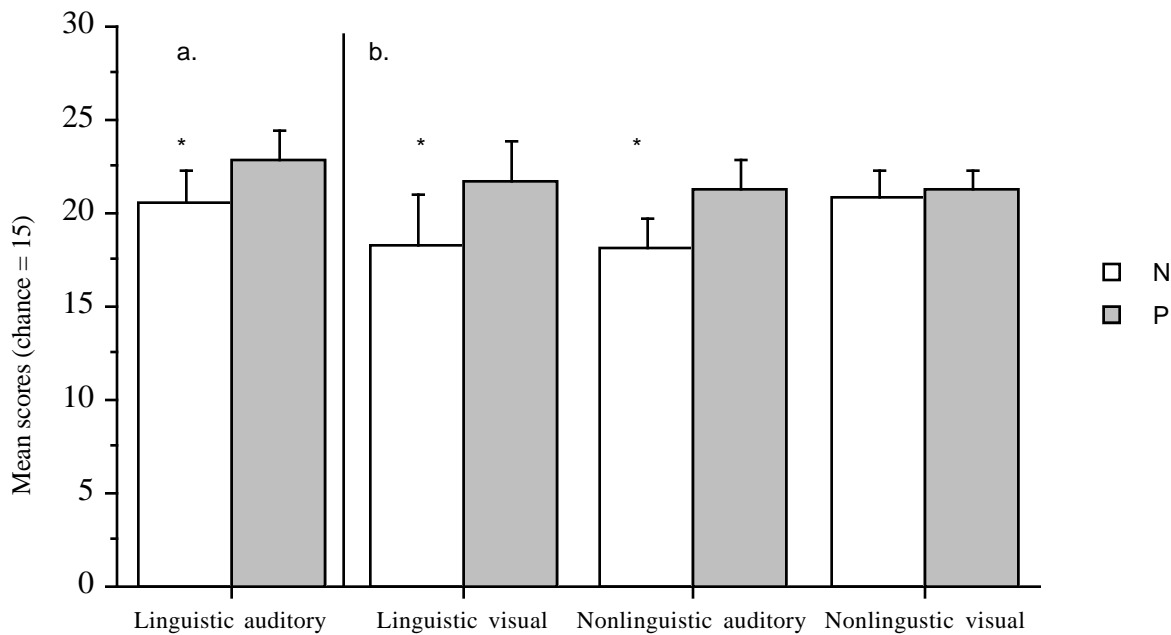
Figure 1: a. Mean scores from Experiment 1. b. Mean scores from Experiment 2.

differences in learning as a function of structural differences between Language P and Language N. To address this question, the scores for the two language groups for items testing each of the five rules were submitted to an ANOVA. The main effect of Language P versus N was significant for the Nonlinguistic auditory [$F(1, 42) = 7.72$, $p < .01$] and the Linguistic visual condition [$F(1, 28) = 4.56$, $p < .05$], but not for the Nonlinguistic visual condition [$F(1, 38) = .23$, n.s.].

In order to ask whether the linguistic or non-linguistic status of the input influenced performance differentially as a function of the availability of linguistic dependencies, we performed a two-way between-subjects ANOVA contrasting Language (P versus N) and Linguistic Status (language versus non-language materials), including the auditory linguistic data from Experiment 1. There was a significant main effect of Language: $F(1, 150) = 15.17$, $p < .0001$. Neither the main effect of Linguistic Status [$F(1, 150) = 1.09$, n.s.] nor the interaction between Language and Linguistic Status [$F(1, 150) = .71$, n.s.] were significant. These analyses indicate that the linguistic status of the input – that is, whether the grammars were implemented in linguistic or non-linguistic tokens – did not affect overall performance. Instead, the dominant factor was whether the input was derived from Language P, which contained predictive dependencies as a statistical cue to phrase structure, or Language N, which did not. This overall non-effect of linguistic status occurred despite the fact that performance on the visual non-linguistic task did not show the predicted difference between Language P and N (see Figure 1). We are currently testing hypotheses concerning why the visual nonlinguistic task patterned differently from the other three conditions included in Experiments 1 and 2.

General Discussion

These studies ask whether predictive dependencies serve a learnability function in the acquisition of language. The results of Experiment 1 suggest that adult learners are better able to acquire an artificial language which contains predictive dependencies as a cue to phrase structure than a comparable language which does not. Experiment 2 extends these results to demonstrate that the use of predictive dependencies in learning phrase structure is not limited to language learning tasks. These findings mirror prior results suggesting that transitional probability computation in word segmentation tasks can occur when 'words' are created from non-linguistic tones (Saffran, Johnson, Newport, & Aslin, 1999) or visuomotor sequences (Hunt & Aslin, 1998).

Predictive dependencies are a hallmark of natural languages. However, it is of interest to note that these general organizational principles are by no means unique to language. Lashley (1951) observed that hierarchical organization characterizes an enormous variety of behaviors: "the coordination of leg movements in insects, the song of birds, the control of trotting and pacing in a gaited horse, the rat running the maze, the architect designing a house, and the carpenter sawing a board present a problem of sequences of action which cannot be explained in terms of successions of external stimuli" (p. 113). Such observations suggest that learners may be biased to process information in a particular fashion, enabling a learning process which results in phrases and hierarchically structured representations.

The kinds of structure at issue here serve to organize and package serial information into manageable chunks, which then enter relationships with one another. This process presumably maximizes cognitive economy, facilitating the

transmission of more complex information than could be transmitted otherwise. Pinker and Bloom (1990) argue that "hierarchical organization characterizes many neural systems, perhaps any system, that we would want to call complex...Hierarchy and seriality are so useful that for all we know they may have evolved many times in neural systems" (p. 726). When applied to syntax, this kind of argument suggests that grammars look the way they do because these kinds of organizational principles are the human engineering solution to the problem of serial order.

It is conceivable that this type of packaging of serial inputs into higher-order organization facilitates not only language production and processing, but also language acquisition. Systems which are highly organized are more learnable than systems which are not -- as long as the system of organization is consistent with the learner's cognitive structure. We anticipate that future research will be extremely useful in further clarifying the extent to which the constraints observed during the process of language acquisition subserve other learning processes as well.

With respect to linguistic structure, one potential theoretical implication of this research concerns an alternative to the traditional innate universal grammar explanation for the pervasiveness of particular linguistic features cross-linguistically.. If human learners are constrained to preferentially acquire certain types of structures, then some of the universal structures of natural languages may have been shaped by these constraints (see also, e.g., Bever, 1970; Christiansen, 1994; Christiansen & Devlin, 1997; Newport, 1982, 1990). Perhaps languages fit our learning abilities so neatly precisely because languages have no choice. If the pertinent learning mechanisms preceded the advent of languages, then there must have been intense pressure for languages to be learnable, with learnability dictated by the structure of human learning mechanisms. On this view, languages evolve to fit the human learner. To the extent that this type of view is correct, then the striking similarities of human languages may be in part the direct reflections of constraints on human learning abilities.

The present research begins the task of recharacterizing language universals in terms of constraints on learning by recasting the distributional features and dependencies inherent in hierarchical phrase structure into cues detected during the learning process. In the case of the constraint to interpret predictive relations as signaling a linguistic unit, the phrase, we find the beginnings of an explanation for why languages ubiquitously contain the within-phrase dependencies initially characterized by structural linguists. Future research will continue to pursue the hypothesis that constraints on learning play an important role in shaping the structure of natural languages. For example, recent computational research suggests that universal word order typologies may in fact reflect the ease with which different types of systems are learned (Christiansen & Devlin, 1997).

With respect to statistical learning, the present research runs counter to the assumption that statistical language learning accounts -- and any other type of theory which assigns an important role to linguistic input -- are necessarily underconstrained. As animal research has amply demonstrated, learning in biological systems is highly constrained (e.g., Garcia & Koelling, 1966; Marler, 1991). There is every reason to believe that statistical learning is similarly constrained; the purported intractability of statistical learning need not be asserted prima facie. What exactly these constraints will turn out to be, and whether they will confer sufficient explanatory power, remain empirical questions. Nevertheless, there are grounds for optimism. Learners are not, and never have been, blank slates. The more we learn about the mechanisms engraved upon that slate, the more we learn about learning.

Acknowledgments

References

Aslin, R. N., Woodward, J. Z., LaMendola, N. P., & Bever, T. G. (1996). Models of word segmentation in maternal speech to infants. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax.* Hillsdale, NJ: Erlbaum.

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

Bever, T. (1970). The cognitive basis for linguistic structures. In J. Hayes (Ed.), *Cognition and the development of language.* New York: Wiley.

Bloomfield, L. (1933). *Language.* New York: Henry Holt.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61,* 93-125.

Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology, 33,* 111-153.

Cartwright, T. A., & Brent, M. R. (1997). Early acquisition of syntactic categories: A formal model. *Cognition, 63,* 121-170.

Christiansen, M. H. (1994*). Infinite languages, finite minds: Connectionism, learning and linguistic structure.* Unpublished Ph.D. dissertation, University of Edinburgh.

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

Christiansen, M. H., & Devlin, J. T. (1997). Recursive inconsistencies are hard to learn: A connectionist perspective on universal word order correlations. *In Proceedings of the Nineteenth Annual Meeting of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Finch, S. P., & Chater, N. (1994). Distributional bootstrapping: From word class to proto-sentence. *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Garcia, J. & Koelling, R. A. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science, 4,* 123-124.

Gleitman, L. R., & Wanner, E. (1982). Language acquisition: The state of the state of the art. In E. Wanner and L. R. Gleitman (Eds.), *Language acquisition: The state of the art.* Cambridge: Cambridge University Press.

Goldowsky, B. (1995). *Learning structured systems from imperfect information.* Unpublished Ph.D. dissertation, University of Rochester.

Goodsitt, J. V., Morgan, J. L., & Kuhl, P. K. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language, 20,* 229-252.

Harris, Z. S. (1951). *Methods in structural linguistics.* Chicago: University of Chicago Press.

Hunt, R. H., & Aslin, R. N. (1998). Statistical learning of visuomotor sequences: Implicit acquisition of subpatterns. In *Proceedings of the Twentieth Annual Meeting of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior: The Hixon Symposium.* New York: Wiley.

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

Marler, P. (1991). The instinct to learn. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition.* Hillsdale, NJ: Erlbaum.

Mintz, T. H. (1996*). The roles of linguistic input and innate mechanisms in children's acquisition of grammatical categories.* Unpublished Ph.D. dissertation, University of Rochester.

Mintz, T. H., Newport, E. L., & Bever, T. G. (1995). Distributional regularities of form class in speech to young children. *Proceedings of NELS 25.* Amherst, MA: GLSA.

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

Morgan, J. L., Meier, R. P., & Newport, E. L. (1989). Facilitating the acquisition of syntax with cross-sentential cues to phrase structure. *Journal of Memory and Language, 28,* 360-374.

Morgan, J. L., & Newport, E. L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior, 20,* 67-85.

Newport, E. L. (1982). Task specificity in language learning? Evidence from speech perception and American Sign Language. In E. Wanner and L. R. Gleitman (Eds.), *Language acquisition: The state of the art.* Cambridge: Cambridge University Press.

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

Perruchet, P., & Vintner, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language, 39,* 246-263.

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

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure.* Cambridge, MA: MIT Press.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences, 13,* 707-784.

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

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

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35,* 606-621.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8,* 101-195.

Schütze, H. (1994). A connectionist model of verb subcategorization. *Proceedings of the 16th Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science, 23,* 569-588.