Journal of Memory and Language **44,** 493–515 (2001)
doi:10.1006/jmla.2000.2759, available online at http://www.academicpress.com on IDEAL®

The Use of Predictive Dependencies in Language Learning

Jenny R. Saffran

University of Wisconsin–Madison

To what extent is linguistic structure learnable from statistical information in the input? This research investigated the role played by statistical learning in the acquisition of rudimentary phrase structure. One type of statistical cue which might assist in the discovery of hierarchical phrase structure given serially presented input is the dependencies, or predictive relationships, present between form classes within phrases. In order to determine whether learners can use this statistical information, adult and child participants were exposed to an artificial language which contained predictive dependencies as a cue to phrase structure. The results suggest that humans possess statistical learning mechanisms which may assist in the acquisition of this abstract component of natural language. ©Academic Press 2001

Key Words: language learning; phrase structure; statistical learning.

Research in nonlinguistic domains suggests that humans are adept at detecting the statistical relationships that characterize the environment. For example, we acquire information about event frequency across a broad range of natural and experimental situations and maintain that information even when there is no reason to do so (Hasher & Zacks, 1984; Hasher, Zacks, Rose, & Sanft, 1987). Such abilities to detect basic statistical properties of the environment are not limited to adults. For example, infants as young as two months of age can detect and remember the contingencies between their own motor movements and a salient environmental event (e.g., Rovee-Collier, 1991). Three-month-old infants can discern the predictive structure in sequences of visual stimuli, showing shorter fixations for predictable new stimuli (e.g., Canfield & Haith, 1991), and 10-month-old infants can learn artificial categories defined only by correlations between features (e.g., Younger, 1985).

Despite these impressive computational abilities, the relationship between statistical learning abilities and the problems confronting language learners is tenuous at best. For statistical properties of language to be useful in language acquisition, the information observable in the input must correspond to the structural features of human languages. Recent computational research suggests that at least for some aspects of language, linguistic structure is mirrored by statistical cues in the input. For example, computational algorithms can use the co-occurrence environments of words in large corpora to discover form classes (e.g., Cartwright & Brent, 1997; Mintz, Newport, & Bever, 1995; Redington, Chater, & Finch, 1998). 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; Per-

This research was supported by NIH Training Grant 5T32DC0003 to the University of Rochester and by NIH Grant DC00167 to Elissa Newport, and the preparation of this manuscript was supported by grants from NIH (HD03352) and the University of Wisconsin Graduate School to JRS. The experiments were part of a dissertation submitted to the Department of Brain and Cognitive Sciences at the University of Rochester. I am grateful to Carlene Abbadecola, Vilma Allejandro, Peter Assad, Sandra Barrueco, Maria Boardman, Toby Calandra, Elizabeth Johnson, Michael Kim, and Lana Nenide for their assistance in conducting these experiments, to Mike Tanenhaus and Greg Carlson for their participation on my dissertation committee, to Morton Christiansen, Michael Maratsos, and three anonymous reviewers for helpful comments on a previous draft, and particularly to Elissa Newport and Dick Aslin for invaluable discussions throughout all stages of this research.

Address correspondence and reprint requests to Jenny R. Saffran, Department of Psychology, University of Wisconsin—Madison, Madison, WI 53706. E-mail: jsaffran@facstaff.wisc.edu.

0749-596X/01 $35.00

ruchet & Vintner, 1998), and behavioral studies suggest that humans, including infants, can detect and use this statistical information (e.g., Aslin, Saffran, & Newport, 1998; Goodsitt, Morgan, & Kuhl, 1993; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996).

While statistical cues may be useful for aspects of language which are tied to the surface properties of the input, such as word segmentation, many other important properties of language are abstract. These aspects of linguistic structure may not be obviously mirrored by the surface structure of the input and thus might not be discoverable by a statistical learning device (e.g., Pinker, 1984). To the extent that this is the case, discovery procedures based on statistical learning will fail. It is thus of interest to ask whether statistical information is available to subserve the acquisition of abstract aspects of language, notably syntax. While there are multiple proposals pertaining to the acquisition of syntactic categories via statistical information (e.g., Billman, 1989; Cartwright & Brent, 1997; Maratsos & Chalkley, 1980; Mintz, 1996; Redington et al., 1998), it remains unclear whether the input contains sufficient learnable statistical information to point learners toward the abstract syntactic structure relating these categories.

Hierarchical Phrase Structure

One abstract feature of human language which is commonly observed cross-linguistically is nonlinear organization. Although words occur serially, our representations of sentences consist of phrases organized into hierarchical relationships, rather than flat, structureless, strings of words. This suggests an interesting learning problem: given serially ordered strings of words as input, how does hierarchical phrase structure arise in learners' representations?

Phrase structure refers to the groupings of categories of words into constituents, which may then themselves enter into new constituents, thereby generating hierarchically organized groupings of elements. For example, the words in the sentence "*The rhino sipped the champagne*" fall into particular groupings: *(The rhino) (sipped (the champagne))*. It would be extremely unnatural to group those words as: *(The) (rhino sipped the) (champagne)*. The correct groupings reflect the phrase structure of English.

A key diagnostic for determining which word sequences cohere as phrases is the distributional behavior of grammatical categories. Phrases are marked by dependencies: a determiner such as *the* requires a noun (forming a noun phrase), and a transitive verb requires an object noun phrase (forming a verb phrase). Phrases are also distributionally highlighted because the words that make up phrases move together within and across sentences. Phrases then interact to generate hierarchical structure. For example, the link between *rhino* and *champagne* in the example sentence above is extremely indirect. *Rhino*, as part of the subject noun phrase, is related to the verb phrase, which in turn contains the noun phrase which contains *champagne*. Other links are far more direct; for example, *rhino* is tightly linked to the determiner *the*. These nonuniform links carry the hierarchical structure of the sentence.

While phrase structure is among the hallmarks of natural languages, it is of interest to note that hierarchical organization is not 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 . . . 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). Hierarchical structure may be present across domains because systems which are highly organized are more learnable, and easier to produce and process, than systems which are not—as long as the system of organization is consistent with the user's cognitive structure. Such considerations suggest that the way that phrase structure works may facilitate its acquisition by language learners.

Morgan and Newport have argued that languages possess cues that serve just this purpose (Morgan, Meier, & Newport, 1987, 1989; Morgan & Newport, 1981). Events at phrase boundaries, such as prosodic cues and functors, and events underlining the unity of entities within the phrase, such as concord morphology and semantic structure, delimit the analyses that learners must perform to discover syntactic structure. On this view, nonsyntactic information correlated with syntactic structure serves to bracket the input into phrases, facilitating learning.

Results from a series of experiments suggest that learners can only use distributional information, such as dependencies, to acquire syntax when additional correlated cues (e.g., semantic or prosodic information) are available to delimit the necessary statistical analyses (Morgan et al., 1987, 1989; Morgan & Newport, 1981). The fact that learners acquire phrase structure only under certain conditions points to the types of constraints which learners bring to bear on linguistic input. On this view, natural languages may contain phrase bracketing devices such as prosody, function words, and concord morphology in part because they facilitate language acquisition (see also Kelly & Martin, 1994). When bracketing information is unavailable, the within-phrase dependencies that carry phrase structure may remain elusive.

This conclusion raises an interesting possibility: perhaps dependencies *themselves* serve as a statistical cue for the discovery of phrase boundaries. Dependencies remain unexplored as a cue for the discovery of phrasal units which may be otherwise unmarked in the input. If dependencies, as purely statistical cues, facilitate acquisition, then perhaps learnability considerations may explain why languages possess this type of structural organization.

The configurational dependencies that underlie phrases have their roots in descriptions by structural linguists (e.g., Bloomfield, 1933; Harris, 1951) and more recently have been codified in the X-bar theory of phrase structure, which reflects commonalities of organization across phrase types (e.g., Jackendoff, 1977). But these dependencies may also be defined statistically, in terms of conditional probabilities computed between form classes: given *Y*, what is the likelihood of *X*? In English, the probability of a simple determiner such as *the* or *a* given a subsequent noun is moderate. But given a simple determiner, the probability that a noun will occur later in the sentence is near unity.

How might dependencies serve as a cue to phrasal units, at least in those languages which contain dependency-defined phrases? If one form class never occurs without another, then this predictiveness signals that the two classes are linked. This connection may be represented as membership in a single unit, such as a phrase. Because form classes can predict one another without being immediately adjacent, a learner attuned to predictive dependencies could, in principle, detect relationships between form classes when other material intervenes (as is typical in natural languages). Thus, the availability of dependencies in the input might lead learners to group form classes together into phrases, even in the absence of other cues correlated with phrase boundaries.

The initial discovery of the phrase has several important ramifications. The first is that phrases serve as a foundation for representations that extend beyond the serial nature of the input. The second related effect is that if one phrase type is available, other dependencies emerge. For example, once noun phrases are discovered, the relationship between transitive verbs and object noun phrases, and prepositions and noun phrases, becomes available. Interestingly, these dependencies are unidirectional. While simple determiners such as *a* and *the* require nouns, nouns do not require determiners. Similarly, prepositions and transitive verbs require object noun phrases, but noun phrases require neither prepositions nor transitive verbs. Note that by focusing on predictiveness rather than simple co-occurrence, spurious phrasal units may be avoided. For example, nearly all English sentences contain subject nouns and verbs. This might lead a learner to incorrectly posit that nouns and verbs are linked as a phrase if co-occurrence is the relevant metric. However, for a learner concerned with predictiveness, all word types equally predict the occurrence of subject nouns and verbs in sentences, since all sentences contain those items. This differs from phrasal dependencies in which certain word types (e.g., determiners) predict other word types (e.g., nouns). We hypothesize that the critical dependencies emerge phrase-internally, where quite specific predictive relationships are available for detection by learners attuned to covariance.

Predictive Dependencies and Other Cues for Sequence Learning

In order for predictive dependencies to be useful for the acquisition of phrase structure, learners must be able to detect these cues amidst myriad additional information in the input, including other surface statistical cues which are less relevant to grammatical structure. Notably, predictive dependencies are not transparently mirrored in the input; to use this type of cue, the learner must be able to discover form classes and then discern the patterns of form classes that demarcate phrasal units. A central debate in the literature on implicit learning processes revolves around just this question: to what extent is performance on artificial language learning tasks driven by surface string and substring cues versus abstract knowledge of the rules underlying the exposure strings? The results of numerous implicit learning studies suggest that learners can utilize the following surface cues in judging grammaticality of novel test sentences: bigram and trigram frequencies, or chunk strength (e.g., Knowlton & Squire, 1996; Perruchet & Pacteau, 1990; Redington & Chater, 1996; E. Servan-Schreiber & Anderson, 1990), frequencies of beginning and ending bigrams and trigrams, or anchor strength (e.g., Perruchet, 1994; Reber & Lewis, 1977), legality of the first element (e.g., Reber & Allen, 1978; Tunney & Altmann, 1999), presence of unique chunks (e.g., Muelemans & Van der Leden, 1997), location of familiar chunks (Gómez & Schvaneveldt, 1994), repetition of items within strings (e.g., Gómez, Gerken, & Schvaneveldt, 1999; Whittlesea & Dorken, 1993), and overall similarity to individual exposure strings (e.g., Vokey & Brooks, 1992).

To what extent might such cues be useful in acquiring phrase structure? Unlike predictive dependencies, these cues are tied tightly to the surface features of the input, more akin to word segmentation than to syntax (e.g., Gómez, 1997; Reber, 1993). This differs from linguistic phrase structure, in which relationships exist between word classes rather than word tokens. Do predictive dependencies allow learners to acquire abstract structure beyond the surface *n*-gram frequency statistics explored in prior artificial grammar learning research?

Prior to addressing the issue of abstract versus surface properties experimentally, it is important to clarify what is meant by the acquisition of abstract structure in language learning. In the implicit learning literature, this type of knowledge is often measured by the extent to which learners can "abstract away from the specific vocabulary used in the training set" (Redington & Chater, 1996, p. 124) to recognize the underlying structure of the input. The primary source of evidence for this type of abstraction has involved transfer studies, in which the degree to which learners have acquired structure beyond surface properties of strings is measured using test items in a new vocabulary (e.g., Altmann, Dienes, & Goode, 1995; Gómez & Schvaneveldt, 1994; Reber, 1969; Shanks, Johnstone, & Staggs, 1997; though see Tunney & Altmann, 1999). However, our underlying knowledge of our native language could not be tapped by a transfer task. Instead, natural language knowledge is related to surface properties in a different way. Words belong to form classes such as nouns and verbs, and regularities over those categories are central to natural language knowledge. Speakers of natural languages can infer the category memberships of novel words only when they occur amidst familiar words, with the exception of prosodic and positional cues which are probabilistically related to category membership (e.g., Kelly, 1992). For this reason, it is unlikely that English speakers could use their knowledge of English to perform grammaticality judgments on sentences implemented in a completely novel vocabulary. Thus, we use the terms *abstract grammatical knowledge* and *rules* to pertain to the types of generalizations characteristic of natural language knowledge[1], which concern the privileges of co-occurrence of form classes, rather than the noncategorical learning typically assessed in transfer tasks.

[1] We use the term *rules* here as a notational convenience to describe the structure of the artificial language used in these experiments; however, we do not claim that the set of rules used here exhausts the possible descriptions of the input or that human linguistic knowledge is represented as symbolic rules.

The first experiment was designed to test the hypothesis that learners are sensitive to the predictive dependencies between form classes that are available to signal phrase structure in language input. The exposure language contained no other cues to phrase structure other than predictive dependencies. Despite the lack of correlated nonstatistical cues, learners attuned to predictive structure should be able to use predictive dependencies to discern phrasal units. At the same time, learners should not be misled by other unrelated surface features of the input corpus.

EXPERIMENT 1

Adult participants were exposed to sentences from an artificial language adapted from the grammar used by Morgan and Newport (1981). Phrase structure rules governed form classes of words. The syntax of words in each category pertained to the distribution of those words with respect to the other categories, but not to any semantic features. Unlike prior studies (e.g., Morgan & Newport, 1981), no visual referents were available to assist learners in determining which forms belong to which categories. Form classes were organized into phrases: for example, A phrases consisted of an A word plus an optional D word. The only cue to grammatical structure was statistical information; no prosodic, semantic, or referential cues were available to supplement the predictive dependencies reflected in the patternings of words.

One potentially important difference from the language used by Morgan and Newport (1981) concerns predictiveness within phrases. The original language was inconsistent in its use of the predictive dependencies which are characteristic of natural languages: a D word could occur either with an A word in an A phrase, or with a C word in a C phrase. We hypothesized that participants in these experiments might have been hindered by the lack of predictive dependencies in the input, thus requiring additional cues correlated with phrase structure. The present language thus contained consistent predictive dependencies as a cue to phrasal units.

The predictive relationships hypothesized to be pertinent for the discovery of phrases were carried by adjacent form classes. This is far simpler than is typically observed in natural languages. Nevertheless, this type of miniature language is sufficiently complex that adult participants in prior studies learned only its simplest structures unless additional cues were available (Morgan & Newport, 1981; Morgan et al., 1987, 1989).

The primary question addressed in this study was the extent to which learners could acquire the language when the only cues to phrase structure were predictive dependencies. A number of other statistical cues previously shown to affect performance on artificial grammar learning tasks in implicit learning studies were also present in the input, but did not serve as cues to phrase structure (e.g., chunk strength, anchor strength, unique pairs, legality of starting elements). Some of these cues are necessarily correlated with grammaticality. For example, an ungrammatical sentence violating the rule "sentences can only begin with a single A word" by beginning with two A words will create a unique pair, AA, not present in the exposure corpus. To tease apart the contributions of grammaticality and surface cues such as frequency and similarity, we performed item analyses of covariance, as reported below.

An additional variable concerned the learning procedure. Half of the participants were explicitly instructed to acquire the grammatical rules of the nonsense language (intentional condition). The other half of the participants were assigned to the incidental condition, in which the primary task was coloring on the computer, with the artificial language presented as a background stimulus (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). This paradigm enforces passive exposure rather than active hypothesis testing. The incidental learning style has been argued to be best suited to language learning (e.g., Reber, 1993; Saffran et al., 1997). The present experiment allowed us to compare the outcomes of incidental and intentional learning and to ascertain whether the mechanisms subserving this type of statistical learning can operate in a mode closer to the exposure-driven learning observed during language development.

Method

Participants

Twenty-nine monolingual English speaking undergraduates at the University of Rochester participated in this study. Fourteen participants were assigned to the intentional condition, and 15 were assigned to the incidental condition. Four additional participants completed only the first session of the experiment; their data were not included in the analysis. Sixteen control participants (10 from the University of Rochester, 6 from the University of Wisconsin–Madison) were tested without any exposure to the language. All participants gave informed consent prior to participating.

Description of the Linguistic System

The artificial language was closely adapted from the language used by Morgan and Newport (1981) and is generated by the rewrite rules in (1):

$$
\begin{aligned}
(1)\quad & S \longrightarrow AP + BP + (CP) \\
& AP \longrightarrow A + (D) \\
& BP \rightarrow \begin{Bmatrix} CP + F \\ E \end{Bmatrix} \\
& CP \longrightarrow C + (G)
\end{aligned}
$$

Each letter in the grammar represents one form class, consisting of two to four monosyllabic nonsense words (see Table 1). Note that D and G are optional, as is the final C phrase; the B phrase has two variants. This grammar generates 18 possible sentence types; only sentences of five or fewer words (14 sentence types) were used. Fifty sentences were randomly chosen as the presentation set, out of the 1624 sentences of five words or fewer. A trained female speaker recorded the presentation set in two random orders with uniformly descending prosody across each sentence. Words were spoken at a rate of approximately one word per second. Approximately three seconds of silence separated each sentence. The recorded block of 100 sentences lasted 7 min.

This language contains the type of predictive structure found in natural languages. In A phrases, A words can occur without D words, but occurrences of D words perfectly predict the presence of A words; the same relationship obtains between C words and G words. Similarly, C phrases can occur without F words (as optional units at the ends of sentences), but if an F word is present, a C phrase must precede it. The directionality of the statistical patterns in this language is the opposite of English, in which perfect predictors precede the member of the phrase that they predict (e.g., determiners precede nouns, prepositions precede noun phrases, and transitive verbs precede their objects). Any attempt to project English structure onto the language would have resulted in poor learning outcomes.

TABLE 1

Word Categories from the Artificial Language

Category A	biff	hep	mib	rud
Category C	cav	lum	neb	sig
Category D	klor	pell		
Category E	jux	vot		
Category F	dupp	loke		
Category G	tiz	pilk		

Procedure

Participants in the intentional condition were told that they would hear a nonsense language consisting of meaningless words arranged into sentences via grammatical rules and that they would be tested on its grammatical structure. Participants in the incidental condition were asked to create an illustration using the children's computer coloring game KidPix2. These participants were informed that there would be a nonsense language playing in the background, but told nothing about the structure of the language. They were also informed that they would be tested on the nonsense language, but not told which aspects of the language would be tested. Because participants knew they would be tested, this condition was not fully incidental. All participants were tested individually.

Each participant heard the tape of sentences four times during a 30-min session on each of two consecutive days, with a break halfway through each session. Three different forced-choice tests, adapted from Morgan and Newport (1981), were administered. On all tests, participants were instructed to circle 1 or 2 depending on whether the first or the second sequence on each trial was a more acceptable string from the exposure language. Participants received several practice trials in English prior to each test.

Rule tests one and two. Rule Test One was administered after the first listening session, and Rule Test Two was administered after the second listening session. These forced-choice tests assessed knowledge of the generalizations over form classes which generated the input. There were 24 item pairs on each test, four testing each of six different rules of the language (see Table 2 for the rules and example test items). On each trial, participants heard a pair of novel sentences, recorded by a trained speaker with uniformly descending intonation. One member of each pair was a grammatical sentence; the other sentence violated a rule of the language. Participants were instructed to choose the grammatical sentence. As shown in Table 2, the first two rules tested whether participants had learned about the privileges of occurrence of individual word classes. The last four rules concerned the dependencies between word classes which signal the phrase structure of the language. The two tests contained different items, as noted in the Appendix, and their order was the same for all participants.

TABLE 2

The Six Rules Tested and Example Test Items

Rule 1: Every sentence must contain at least one A word.	
MIB SIG DUPP	[A - C - F]
*SIG DUPP	[C - F]
Rule 2: No sentence may contain more than one A word.	
MIB PELL JUX CAV	[A - D - E - C]
[a] MIB BIFF PELL JUX CAV	[A - A - D - E - C]
Rule 3: If there is an E word, there cannot be a CP.	
BIFF KLOR JUX	[A - D - E]
[a] BIFF KLOR LUM JUX	[A - D- C - E]
Rule 4: If there is a D word, then there must be an A word.	
RUD PELL NEB DUPP SIG	[A - D - C - F - C]
[a] PELL NEB DUPP SIG	[D - C - F - C]
Rule 5: If there is an F word, then there must be a CP.	
BIFF NEB DUPP	[A - C - F]
[a] BIFF DUPP	[A - F]
Rule 6: If there is a G word, then there must be a C word.	
MIB VOT CAV TIZ	[A - E - C - G]
[a] MIB VOT TIZ	[A - E - G]

[a] Ungrammatical items.

While the rule tests provided information about the extent to which participants acquired the structure of the language, phrase structure knowledge was not required for successful performance. For example, participants could perform better than chance on the items testing Rule 5 by noting that they had never heard an F word directly following an A word, without recourse to a phrase structure representation.

Fragment test. This test was intended to more directly assess the extent to which learners represented the input in terms of phrasal groupings. Each trial consisted of two sentence fragments, a phrase and a sequence spanning a phrase boundary. We hypothesized that if learners had succeeded in grouping the input strings into phrases, then sentence fragments which constituted phrases should appear more natural than nonphrase fragments. For example, the English phrase *the dog* should appear more coherent than the nonphrase fragment *bit the*, even though both word sequences are consistent with English grammar. Participants were thus asked to decide which fragment seemed like a better or more coherent group or unit from the nonsense language. To ensure that performance on this test was a function of phrasal knowledge rather than the frequencies with which each fragment had occurred in the input, the phrase and nonphrase fragments were controlled such that both fragment types were equally frequent in the exposure corpus. If learners attend only to fragment frequency information, they should perform at chance due to the controls on word pair frequencies. Performance exceeding chance would suggest that participants' representations of phrasal fragments were more coherent than their representations of nonphrase fragments, despite the fact that both fragment types occurred equally often in the input. Phrase sequences were coded as grammatical and nonphrase sequences as ungrammatical, although all were legal sequences in the language. The test consisted of 24 items, eight items testing each of the three phrase types; each phrasal category was tested by two different nonphrase fragment types (see Table 3). This test was administered only after the second listening session.

TABLE 3

Sample Items from the Fragment Test

A phrase	BIFF KLOR	[A - D]
	[a] BIFF CAV	[A - C]
	[a] KLOR CAV	[D - C]
B phrase	SIG TIZ LOKE	[C - G - F]
	[a] PELL SIG TIZ	[D - C - G]
	[a] TIZ LOKE CAV	[G - F - C]
C phrase	SIG PILK	[C - G]
	[a] KLOR SIG	[D - C]
	[a] DUPP SIG	[F - C]

[a] A word pair crossing a phrase boundary rather than ungrammaticality.

Control group. An additional group of participants received the three tests without exposure to the language to ensure that performance exceeding chance by the experimental participants was due to learning rather than biased test materials (e.g., artifacts of similarity to English).

Results

The first analysis tested differences between the three groups: the intentional condition, the incidental condition, and the control group. An ANOVA compared overall scores on the three tests. The main effect of Group (intentional, incidental, and control) was significant: $F(2,42) = 19.3$, $p < .0001$. A two-tailed Dunnett test showed that this effect was due to significantly better performance in the two experimental groups as compared to the control group (both $p < .01$). No differences between the experimental groups emerged. There was also a significant main effect of Test (Rule Test 1, Rule Test 2, and Fragment Test): $F(2,42) = 7.3$, $p < .01$. This effect was due to better performance on both rule tests than on the Fragment Test (both $p < .01$). No interaction between Group and Test emerged. Because the two experimental groups performed equivalently on all three tests, data from the intentional and incidental conditions are combined in the subsequent analyses.

The next set of analyses contrasted the experimental and control groups' performance on the three tests. Experimental participants significantly outperformed control participants on all three tests; Rule Test One: $t(43) = 3.33$, $p < .001$; Rule Test Two: $t(43) = 5.79$, $p < .0001$; Fragment Test: $t(43) = 2.96$, $p < .001$ (see Fig. 1).

Table 4 presents participants' mean scores on the six rules tested on the two rule tests, along with significance tests contrasting the experimental group to chance and to the control group, a more conservative measure of performance (Redington & Chater, 1996). Experimental participants significantly exceeded chance on four rules on Rule Test One, with performance significantly below chance on one rule (discussed below), and exceeded chance on five rules on Rule Test Two. Experimental participants outperformed controls on three Rule Test One rules and four Rule Test Two rules. Successful per-

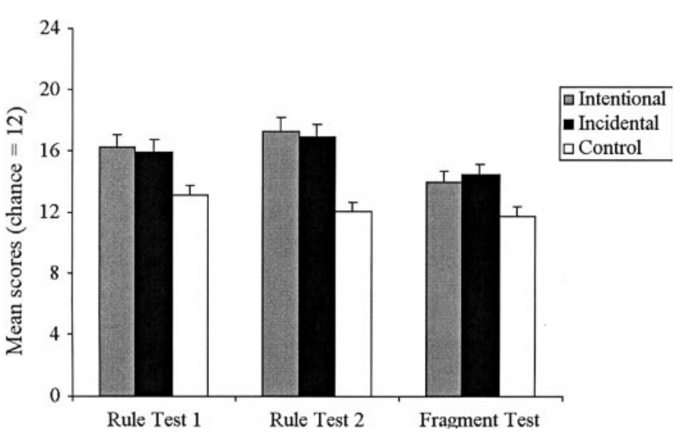

FIG. 1. Mean scores for participants (experimental and control) in Experiment 1.

TABLE 4

Experiment 1: Experimental Group Mean Scores and Significance Tests Against Chance (2 out of 4 Possible) and Against the Control Group for Rule Tests One and Two

	Experimental mean	*t*-test against chance (df = 28)	Control mean	Control vs. experimental (df = 43)
		Rule Test One		
Rule 1	3.31	7.03 **	2.25	3.25 **
Rule 2	2.52	2.56 *	1.68	2.14 *
Rule 3	1.17	–5.53 **	1.56	1.26
Rule 4	3.48	10.18 **	2.5	3.48 **
Rule 5	3.03	6.15 **	2.75	.99
Rule 6	2.55	3.13 *	2.38	.54
		Rule Test Two		
Rule 1	3.10	5.06 **	2.18	2.59 *
Rule 2	2.59	2.82 *	2.13	1.34
Rule 3	1.72	–1.62	1.94	.722
Rule 4	3.45	11.4 **	1.94	4.90 **
Rule 5	3.03	8.19 **	2.31	2.59 *
Rule 6	3.14	7.35 **	1.63	4.96 **

*$p < .05$.
**$p < .01$.

formance included the conditional rules governing the dependencies between grammatical categories. On the Fragment Test, the experimental group's scores on items testing the B and C phrases significantly exceeded both chance and the control group's scores, with chance performance on the A phrase items (see Table 5).

We next performed a series of item analyses to determine the basis for participants' grammaticality judgments. As noted above, novel ungrammatical items tend to differ from the exposure corpus in more than just grammaticality, and these surface factors have been shown to influence participants' performance in artificial grammar learning tasks. To determine which factors influenced participants' judgments, we performed analyses of covariance (ANCOVA) in which string and substring features were entered as covariates. The question of interest was whether grammaticality (whether or not a given test item violated a rule of the language) would continue to account for a significant portion of the variance once other factors representing surface characteristics of the stimuli were entered into the model.

Each test consisted of 24 forced-choice pairs contrasting grammatical and ungrammatical items, rendering 48 items for each ANCOVA. The dependent variable was the proportion of times each item was endorsed as grammatical. Items were then coded according to measures shown to be pertinent in prior artificial grammar learning studies, as described previously in the literature review. For the ANCOVA models examining Rule Tests One and Two, grammaticality was coded as a two-level factor: items were either grammatical or not. Legality of the first word was also coded as a two-level factor. The remaining factors were all continuous variables computed for each test item relative to the exposure corpus: chunk strength (the average of the input frequencies for all word pairs and triples for each item), anchor strength (the composite of the input frequencies for the initial and final word pairs and triples for each item),[2] uniqueness (the

[2] Additional analyses tested bigrams and trigrams separately for both the chunk strength and anchor strength covariates, as well as initial versus final anchors. The results did not differ when the sizes and positions of chunks were taken into account.

TABLE 5

Experiment 1: Mean Scores and Significance Tests against Chance (4 out of 8 Possible) and against the Control Group for the Fragment Test

Experimental mean		*t*-test against chance (df = 28)	Control mean	Control vs. experimental (df = 43)
A phrase	4.00	0	3.75	.423
B phrase	5.21	4.74 **	4.31	2.13 *
C phrase	5.14	3.84 **	3.75	2.86 **

*$p < .05$.
**$p < .01$.

number of word pairs in each item that never occurred in the input), and similarity (the number of words by which each item differed from the most similar sentence in the input). In addition, we included the length of each test item as a factor, because the grammatical items were longer than the ungrammatical items for four of the six rules tested.[3] Three additional factors shown to influence judgments in other studies were not relevant to our test stimuli and were not included in the analyses: repetition (test sentences did not contain word repetitions), final word legality (all final words were legal), and chunks in impermissible locations (only four items contained chunks in impermissible locations).

Because the Fragment Test items were not full sentences, the fragment item analyses contained only a subset of the variables described above: grammaticality (phrase vs. nonphrase), chunk strength, uniqueness, and similarity. An additional variable concerned violations of predictive dependencies. Recall that each phrase fragment was tested by two different types of nonphrase fragments. Some of the nonphrase fragments (e.g., DC) violated one or more predictive dependencies (e.g., D predicts A). Other nonphrase fragments, such as AC, did not violate any predictive dependencies because neither A nor C requires any additional element within the phrase. None of the phrases (grammatical fragments) violated predictive dependencies. We were interested in asking whether the degree to which fragments violated predictiveness would influence participants' judgments. Note that while violations of predictive dependencies are related to grammaticality (phrase versus nonphrase) in that the presence of a violation makes a fragment ungrammatical, the two variables are not totally overlapping; fragments that are not phrases did not always violate predictive dependencies.

An underlying assumption of ANCOVA is homogeneity of regression slopes. To test this assumption, we first examined the interaction effects between the two factors and each of the covariates for all three tests. None of the interactions were significant, consistent with homogeneity of regression slopes. Because the assumption of homogeneity of slopes cannot be rejected, the effects of the covariates can be estimated by a single slope, and the interaction terms were eliminated from the final models.

ANCOVA models were generated to assess the item results from each of the three tests, including both the experimental group and the control group separately as dependent variables.[4] For the experimental group, only grammaticality was a significant predictor of scores on both Rule Tests One and Two; no other covariates accounted for a significant portion of the variance. On Rule Test One, the legality of the first word factor ($p < .07$) and the length of test items factor

[3] Thanks to an anonymous reviewer for this suggestion.

[4] To ensure that effects of covariates were not minimized by the inclusion of multiple potentially related variables, we also performed a series of single-factor plus single-covariate analyses (e.g., grammaticality × chunk strength). The pattern of results corresponded to the omnibus ANCOVA models for all three tests.

($p < .09$) showed a trend toward significance; the length factor also showed a trend toward significance on Rule Test Two ($p < .07$). Attention to the length of test items may explain the pattern of results for Rule 3. For this rule, subjects chose the correct item less often than would be expected by chance, a pattern potentially due to the fact that the correct answer always contained fewer words than the incorrect answer. When the control group means served as the dependent variable, the length of the test items accounted for a significant portion of the variance on Rule Test One; no covariates were significant on Rule Test Two (see Table 6).

Three separate ANCOVA models were applied to the results from the Fragment Test, because correlational analyses indicated that two of the variables, grammaticality and violations of dependencies, were highly related to one another ($r = .65$). When the violations of predictive dependencies covariate was not included in the ANCOVA, grammaticality was a significant predictor. Similarly, the violations factor was significant when grammaticality was not included in the analysis. However, when the violations covariate and the grammaticality factor were both included, only the violations covariate was a significant predictor; the effect of grammaticality was removed. These findings suggest that the number of violations of dependencies in the test fragments affected responses more strongly than whether or not a fragment corresponded to a phrasal unit. No other factors were significant in these analyses. None of the variables were significant when the control group means served as the dependent variable (see Table 7).

TABLE 6

ANCOVA Results for Rule Tests One and Two, Experimental and Control Groups

Factor	Experimental group F-value (df = 1, 40)	Control group F-value (df = 1, 40)
	Rule Test One	
Grammaticality	9.57 **	.76
First word legality	3.56	.06
Chunk strength	.01	.53
Anchor strength	.01	.07
Similarity	.01	.82
Unique pairs	.02	.06
Length	3.05	4.36 *
	Rule Test Two	
Grammaticality	22.52 **	.38
First word legality	2.31	.01
Chunk strength	.15	.84
Anchor strength	.35	1.23
Similarity	.61	.01
Unique pairs	.13	.17
Length	3.54	.04

*$p < .05$.
**$p < .01$.

Discussion

The results suggest that learners can detect phrasal units in the absence of relevant cues other than predictive dependencies. Performance on the rule tests suggest that learners acquired information regarding the occurrence of individual categories as well as the more difficult conditional rules governing dependencies between categories. These findings differ from the results of the studies by Morgan, Newport, and colleagues (Morgan & Newport, 1981; Morgan et al., 1987, 1989) in which learners did not have access to consistent predictive dependencies. The predictive relationships between form classes used in the present experiment apparently facilitated the statistical learning of phrasal groupings. Importantly, the item analyses suggest that even when variance due to other statistical properties of the test strings, such as chunk frequency, similarity, and legality of the first word, is partialled out, grammaticality continues to account for overall performance.

The experimental group also outperformed the control group on the Fragment Test. Phrase and nonphrase test items were equated for frequency in the input corpus. Thus, raw frequency counts alone cannot serve to distinguish phrases from nonphrases on this test. Instead, successful performance suggests that learners' representations superceded linear co-occurrence. The results indicate that participants learned the B and C phrases in this fashion, but not the sentence-initial A phrase.

TABLE 7

ANCOVA Results for the Fragment Test, Experimental and Control Groups, both with and without the Violations of Phrase Structure Factor

Factor	Experimental group *F*-value (df = 1,42)	Control group *F*-value (df = 1,42)
Fragment Test (full analysis)		
Grammaticality	.21	.61
Chunk strength	1.92	.01
Similarity	.83	.02
Unique pairs	.03	.27
Violations of dependencies	5.24 *	.44
Factor	**Experimental group *F*-value (df = 1,43)**	**Control group *F*-value (df = 1,43)**
Fragment Test (excluding violations of predictive dependencies)		
Grammaticality	5.65 *	.22
Chunk strength	1.90	.01
Similarity	.35	.01
Unique pairs	.09	.47
Factor	**Experimental group *F*-value (df = 1,43)**	**Control group *F*-value (df = 1,43)**
Fragment Test (excluding grammaticality)		
Chunk strength	3.43	.09
Similarity	.98	.02
Unique pairs	.01	.46
Violations of dependencies	14.23 **	.04

$^*p < .05$.
$^{**}p < .01$.

If learners detected predictive dependencies, as hypothesized, then the extent to which fragments violated dependencies should have influenced the degree to which learners endorsed them as phrases. The item analyses support this hypothesis: the number of predictive dependencies violated by each test fragment was a stronger predictor of participants' judgments than whether or not a fragment was a phrase. Moreover, these considerations suggest a potential reason why participants did not successfully discriminate A phrase fragments from nonphrase fragments. While all of the nonphrase items for the other phrase types violated at least one predictive dependency, half of the nonphrase items testing the A phrase did not violate any predictive dependencies. These findings offer indirect evidence supporting the claim that predictive dependencies are playing a central role in the learning process.

Humans are thus capable of at least the rudimentary acquisition of one aspect of syntactic organization from statistical information in a laboratory learning task. Moreover, participants were able to do so in an incidental paradigm, in which learning was a secondary task. The learning abilities which subserve this process are likely to be deployable automatically, as would be expected of mechanisms underlying child language acquisition. To assess the presence of these learning mechanisms in younger learners, the second experiment extended the investigation of dependency cues to include children.

EXPERIMENT 2

Experiments using complex artificial language learning paradigms have primarily involved only adult participants (though see, e.g., Braine, Brody, Brooks, Sudhalter, Ross, Catalano, & Fisch, 1990; Gómez & Gerken, 1999; Saffran et al., 1997). Children tend not to be included either because researchers are not concerned with the relationship between artificial grammar learning and first language acquisition or due to the difficulties inherent in eliciting metalinguistic judgments from children. However, data from child learners are useful for characterizations of the learning mechanisms available for first language acquisition and for modeling child language learning processes.

Saffran et al. (1997) attempted to equate task demands for adult and child learners in a word segmentation task by using the incidental learning task described in Experiment 1, in which learners were engaged in a cover task of coloring on the computer. Under those circumstances, children and adults showed equivalent levels of performance. We thus tested first- and second-grade children on the incidental grammar learning task employed in Experiment 1 to determine whether child learners can detect and use predictive dependencies in acquiring the beginnings of phrase structure.

Method

Participants

Twenty-six monolingual English-speaking children participated in this study. The children ranged in age from 6 to 9 (mean age: 7 years 7 months) and were recruited from local summer camps. Two additional children were excluded from the analyses because one of their parents spoke a language other than English in the home. Six additional children only completed the first of the two experimental sessions and were not included in the analysis. The children received stickers and a color printout of the computer drawings they produced during the experiment. An additional group of 21 6- to 9-year-old children (mean age: 8 years 3 months) were recruited to serve as control participants for this study; these children received only the three tests.

Procedure

The language, test materials, and incidental learning paradigm were the same as those used in Experiment 1. One additional variable concerned the length of exposure. Pilot data suggested that the two 28-min exposure sessions used in Experiment 1 might be overly lengthy given children's attentional resources. We thus shortened the exposure sessions for 15 of the experimental children by one third to 21 min to lessen the effects of fatigue. In order to maintain the children's interest during testing, they were given a sticker to place on a sticker drawing after every fourth trial. All children received multiple trials in English before each test to ensure that they understood the instructions.

Results

The first analysis contrasted the two listening time groups. As no significant differences emerged between the 21- and 28-min listening session groups, the two groups are combined in the subsequent analyses. The next set of analyses examined children's overall performance on the three tests (see Fig. 2). An ANOVA compared overall scores on the three tests for the children in the experimental and control groups. The main effect of Group (experimental versus control) was significant: $F(1,45) = 12.85$, $p < .001$. Neither the effect of Test nor the interaction between Group and Test were significant ($F < 1$). These findings suggest that the children in the experimental group outperformed the children in the control group overall.

The next set of analyses examined performance on each of the three tests individually. The experimental children's performance was reliably better than chance on all three tests [Rule Test One, $t(25) = 4.03$, $p < .01$; Rule Test Two: $t(25) = 4.91$, $p < .01$; Fragment Test: $t(25) = 2.49$, $p < .05$]. The experimental children's data were

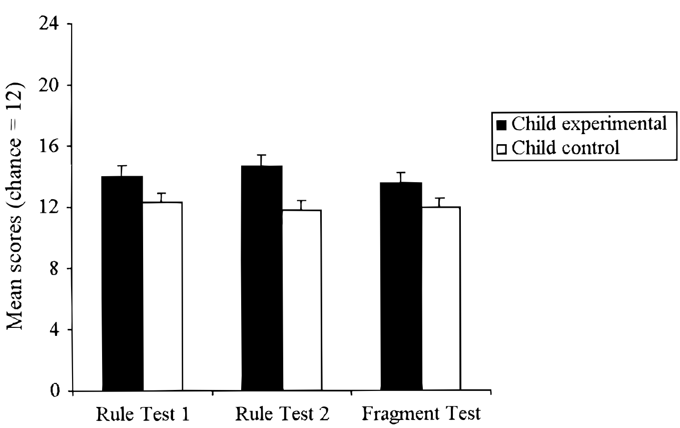

FIG. 2. Mean scores for child participants (experimental and control) in Experiment 2.

then contrasted with the child control participants. The experimental children significantly outperformed the child control participants on Rule Test One [$t(45) = 2.02$, $p < .05$], and Rule Test Two [$t(45) = 3.22$, $p < .01$], with marginal performance on the Fragment Test [$t(45) = 1.87$, $p < .07$].

Table 8 presents participants' mean scores on the six rules tested on the two rule tests, with significance tests contrasting the experimental children to chance and to the control children. The experimental children performed significantly better than would be expected by chance on four rules on both Rule Tests One and Two and performed significantly worse than would be expected by chance on one rule on Rule Test One (discussed later). The experimental children outperformed the control children on three Rule Test One rules, with performance significantly worse than controls on one rule. The experimental children outperformed the control children on two Rule Test Two rules. As with the adult participants, successful performance included the conditional rules testing the acquisition of the dependencies between grammatical categories. On the Fragment Test, the experimental children's scores on the three phrase types exceeded chance only for the B phrase and did not significantly exceed the scores of the child control group (see Table 9).

As in Experiment 1, we performed analyses of covariance to determine the basis for the experimental children's responses (see Table 10). On Rule Test One, only the length of item variable accounted for a significant portion of the variance. On Rule Test Two, both grammaticality and the length of item variable were significant predictors. In the Fragment Test analyses (see Table 11), grammaticality accounted for a significant portion of the variance when the violations of predictive dependencies variable was not included in the model. Similarly, the violations variable accounted for a significant portion of the variance when the grammaticality factor was not included. When the violations variable was included, only a trend toward a significant effect of grammaticality remained ($p < .07$). No other factors made significant contributions to the model. This pattern of results suggests that the experimental children were sensitive to the

TABLE 8

Experiment 2: Experimental Children's Mean Scores and Significance Tests against Chance (2 out of 4 Possible) and against the Control Children for Rule Tests One and Two

	Child experimental mean	*t*-test against chance (df = 25)	Child control mean	Experimental vs. control (df = 45)
		Rule Test One		
Rule 1	2.84	4.90 **	2.19	2.15 *
Rule 2	1.69	–1.03	1.86	–.59
Rule 3	.89	–5.14 **	2.05	–3.12 **
Rule 4	2.89	3.45 **	2.09	2.06 *
Rule 5	2.81	3.25 **	2.24	1.41
Rule 6	2.85	4.28 **	1.86	2.87 **
		Rule Test Two		
Rule 1	3.00	4.82 **	2.05	2.89 **
Rule 2	1.69	–1.55	1.81	–.36
Rule 3	1.58	–1.84	1.91	–.89
Rule 4	2.89	4.52 **	1.95	3.03 **
Rule 5	2.89	3.54 **	2.29	1.63
Rule 6	2.62	2.61 *	1.95	1.70

*$p < .05$.
**$p < .01$.

TABLE 9

Experiment 2: Mean Scores and Significance Tests against Chance (4 out of 8 Possible) and against the Child Control Group for the Fragment Test

	Child experimental mean	*t*-test against chance (df = 25)	Child control mean	Experimental vs. control (df = 45)
A phrase	4.39	1.51	4.00	.32
B phrase	4.69	2.08 *	4.19	.33
C phrase	4.39	1.08	3.67	1.56

*$p < .05$.

phrase structure and presence of predictive dependencies in the input. No variables contributed significantly to the ANCOVA models for the child control participants.

As with the adult participants, the experimental children performed worse than would be expected by chance on Rule 3 on Rule Test One. This pattern of results may be due to the length of the test items; for Rule 3, the correct answer is shorter than the incorrect answer on each trial. The analyses of covariance suggest that length strongly influenced children's performance. Whether this is due to a response bias or due to characteristics of the input is not clear. In the case of the adult control participants, who also showed a significant effect of test item length, effects of length could not have been due to characteristics of the exposure stimuli since they did not hear the exposure stimuli. One possible reason why the experimental child group and the adult control group might show significant effects of length in the item analyses, but not the experimental adult group or the child controls, is that the experimental child participants, who had not yet learned much about the structure of the language, followed a similar strategy to the adult controls: comparing the two items in each test pair and noting that one was missing a word contained in the other. In the absence of good structural knowledge about the language, the full items may appear to be more grammatical than items missing a word. The child control participants may not have been influenced by length because they lacked the task understanding required to do more than randomly guess on each trial, whereas the adult experimental group may have already acquired sufficient structural knowledge to perform the task.

The final set of analyses compared the experimental children's performance with the experimental adult participants in Experiment 1. An ANOVA compared overall scores on the three tests for the two age groups. There was a significant main effect of Age, with adults outperforming children: $F(1,53) = 10.52$, $p < .01$. There was also a significant main effect of Test:

TABLE 10

ANCOVA Results for Rule Tests One and Two, Child Experimental and Child Control Groups

Factor	Child experimental group *F*-value (df = 1,40)	Child control group *F*-value (df = 1,40)
	Rule Test One	
Grammaticality	2.13	.01
First word legality	.28	.39
Chunk strength	.12	.02
Anchor strength	.03	.46
Similarity	1.32	.08
Unique pairs	.02	1.41
Length	12.11 **	.97
	Rule Test Two	
Grammaticality	7.76 **	1.00
First word legality	3.17	.81
Chunk strength	.34	.09
Anchor strength	.17	1.18
Similarity	1.11	.15
Unique pairs	.68	.84
Length	8.58 **	1.40

*$p < .05$.
**$p < .01$.

TABLE 11

ANCOVA Results for the Fragment Test, Child Experimental and Child Control Groups

Factor	Child experimental group *F*-value (df = 1,42)	Child control group *F*-value (df = 1,42)
Fragment Test (full analysis)		
Grammaticality	.21	.01
Chunk strength	1.92	.39
Similarity	.83	.02
Unique pairs	.03	1.95
Violations of dependencies	5.24 *	.09

Factor	Child experimental group *F*-value (df = 1,43)	Child control group *F*-value (df = 1,43)
Fragment Test (excluding violations of predictive dependencies)		
Grammaticality	5.65 *	.02
Chunk strength	1.90	.01
Similarity	.35	.01
Unique pairs	.09	.03

Factor	Child experimental group *F*-value (df = 1,43)	Child control group *F*-value (df = 1,43)
Fragment Test (excluding grammaticality)		
Chunk strength	1.08	.39
Similarity	1.03	.02
Unique pairs	.01	2.01
Violations of dependencies	7.05 *	.01

*$p < .05$.

$F(2,53) = 9.65$, $p < .01$. This effect was due to better performance on both rule tests than on the Fragment Test (both $p < .01$). No interaction between Age and Test emerged.

Discussion

The results of Experiment 2 suggest that children may possess a limited ability to acquire syntactic knowledge via statistical information. While their performance was not as strong as the adults', the children did acquire rudimentary aspects of the phrase structure of the language; this was particularly evident in the analyses of covariance, in which the grammaticality of the test items accounted for a significant proportion of the variance in children's responses. Performance on the conditional rules suggests that the children acquired some of the dependencies of the phrase structure. Child learners may be limited in their ability to detect and utilize predictive dependencies. Alternatively, the difficulty of the metalinguistic judgments required by these testing procedures, particularly on the Fragment Test, may have masked children's linguistic knowledge; young children tend to perform more poorly than older children and adults on psycholinguistic tests (e.g., Fathman, 1975). For example, Slavoff and Johnson (1995) found that children under the age of 7 1/2 were unable to consistently perform grammaticality judgments for native language sentences. Future studies will entail the development of more implicit measures of learning that do not rely on forced-choice judgments. Listening time measures have been profitably applied to the study of infant learning of simple grammars (Gómez & Gerken, 1999; Marcus, Vijayan, Bandi Rao, & Vishton, 1999) and may be adaptable for studying the acquisition of more complex grammars such as those used here.

Age differences did emerge in these results: adults consistently outperformed the children on all three tests. The children nevertheless displayed some systematicity in their responses. As shown in Table 12, we computed the overall percentage of variance (R^2) in participants' responses that can be accounted for by the factors entered into the analyses of covariance (excluding the length factor; because this factor contributed significantly for the control subjects, the use of item length may be a function of testing strategy not relevant to learning during exposure). For both the adults and the children, these factors accounted for a significant portion of the variance for the experimental participants across age groups on all three tests. This was particularly true of Rule Test Two, where these variables accounted for 62% of the variance in adults' performance and 45% of the variance in children's performance. Like the adults, the chil-

TABLE 12

Percentage of Variance (R^2) in Item Scores Accounted for by the ANCOVA Variables (Excluding Item Length) for Adult, Child, and Control Subjects

	Rule Test One (df = 6,41)	Rule Test Two (df = 6,41)	Fragment Test (df = 5,42)
Adult (Experiment 1)	.50 **	.62 **	.31 **
Child (Experiment 2)	.27 *	.45 **	.28 *
Adult control (Experiment 1)	.08	.06	.03
Child control (Experiment 2)	.07	.07	.05

*$p < .05$.
**$p < .01$.

dren were able to exploit some of the systematic structure available in the input. Whether the children's overall poorer performance is due to their overall learning capacity or to the particular task demands of this experiment is unclear. Even in natural language learning, children start out slower than adults in the early stages of learning (e.g., Krashen, Scarcella, & Long, 1982), even though children eventually surpass adult learners (e.g., Johnson & Newport, 1989; Newport, 1990). It is possible that with additional opportunities for learning, the children would outperform the adults on this task; they may require more data upon which to perform the pertinent analyses, given constraints on their information processing capacities (Slavoff & Johnson, 1995). Alternatively, the acquisition of basic phrase structure may be an aspect of language acquisition which does not show critical period affects, akin to the acquisition of basic word order (e.g., Johnson & Newport, 1989; Newport, 1990).

GENERAL DISCUSSION

The statistical structure of languages represents a potential goldmine for learners. Statistical information could be brought to bear on a variety of the learning problems solved by children acquiring languages, particularly when combined with other types of cues available in linguistic input. It is surprising, in light of this potential wealth of information, that the empirical literature has largely neglected the possibility that statistical learning subserves significant aspects of language acquisition (see also Seidenberg, 1997).

The present research supports the hypothesis that there is a relationship between human learning abilities and the statistical cues which mirror aspects of the structural organization of natural languages. In particular, the dependencies that characterize linguistic phrase structure might be detected by a suitably able learner and used to determine groups of words which cluster into phrases. One ramification of this proposed learning process is that dependencies delimit the learner's subsequent analyses such that syntactic relations *within* phrases are highlighted. A second result is that dependencies *between* phrases and other form classes become evident once the learner's representations include initial phrasal groupings.

The experimental results suggest that human learning mechanisms contain design features suited to the kind of solution demanded by the structure of this learning problem. Adults, and to some extent children, acquired the beginnings of phrase structure given only the basic distributional cues inherent in the dependencies between form classes. These results support the hypothesis that human learning mechanisms possess the computational power needed to derive the beginnings of hierarchical structure from the statistical relationships between form classes and can do so via incidental learning mechanisms.

Interestingly, learners' performance was not measurably influenced by features of the test strings other than grammaticality (and, for the children, string length). There were certainly numerous substring statistical properties of the input which could have driven performance;

learners might have attended primarily to the frequencies of word pairs or triplets or to the presence of unique pairs in test strings. However, learners acquired more abstract and unobservable features of the input—which types of words predict other types of words—thereby deriving phrasal units. The persistent effects of grammaticality may be due to a number of factors. One possibility is that learners were forced to acquire abstract properties of the input because the grammar was written over word types rather than word tokens. In the standard artificial grammar task, rules involve relationships between individual tokens (e.g., B is followed by M or V). However, in the language used here, it would have been very difficult for learners to attend to token relationships due to the large number of possible pairwise combinations (e.g., *biff* could be legally followed by *klor*, *pell*, *cav*, *lum*, *neb*, *sig*, *jux*, or *vot*). Other studies have suggested that the size of the input language influences the degree to which learners abstract away from individual string and substring properties (e.g., McAndrews & Moscovitch, 1985; Muelemans & Van der Leden, 1997).

Another factor which may influence the degree to which learners abstract grammatical properties concerns the grammar itself. The present research emerged out of a tradition in which miniature languages containing natural language-like properties serve as a tool to investigate basic processes in language acquisition (e.g., Braine, 1971; Braine et al., 1990; Moeser & Bregman, 1972; Morgan et al., 1987, 1989; Morgan & Newport, 1981; Saffran et al., 1996, 1997; Valian & Coulson, 1988). However, the artificial grammars employed in studies of implicit learning typically contain few, if any, structural properties which play a role in natural language learning. Moreover, studies of implicit learning are less concerned with the manner in which participants in sequence learning tasks discover units (such as phrases) in the input, focusing instead on the acquisition of rules in the form of (*n*th-order) transitions between items in sequence.

One potentially relevant implicit learning model was proposed by E. Servan-Schreiber and Anderson (1990), although it was not intended as a model of natural language acquisition. The theory of competitive chunking suggests that learners are sensitive to chunks in the input and use known chunks to distinguish grammatical and ungrammatical sequences. At one level, this view is very similar to our own: phrases are detected as chunks in the input. The interesting differences arise when we consider the grounds for discovering and representing chunks. In E. Servan-Schreiber and Anderson's (1990) model, the probability of a chunk winning the competition to enter the final representation is a function of how frequently and recently that chunk was used. According to the predictive dependency hypothesis described here, phrases are discoverable when the presence of one element is tightly linked to the other. This sort of distinction, between raw frequency and conditional frequency (probability), has been shown to be pertinent in other learning tasks: for example, Aslin et al. (1998) found that 8-month-old infants relied on the computation of transitional probabilities between syllables, rather than simply detecting syllable–pair frequencies, to segment novel words from continuous speech. With respect to discovering the phrase structure of natural languages, chunk frequency is likely to be more misleading than the predictive cues characteristic of linguistic phrases.

Can the competitive chunking model account for the present data? Because competitive chunking creates units that are tied tightly to the surface features of the input, a chunk might be something like *biff cav*, from the sequence *biff cav lum loke*. This type of information is very useful for tasks in which the grammar is written over the vocabulary itself (e.g., *biff* may be followed by *cav* or *sig*). However, this type of model would have more difficulty when rules are written over categories of vocabulary, as in the present experiment, and in natural languages. Thus, if the model had seen *biff cav lum loke* and *mib sig lum loke*, but had never seen the grammatical string *biff sig lum loke*, the chunk strength of *biff sig* would be relatively low, incorrectly supporting a judgment of ungrammaticality. It would be interesting to see how the model would perform on such tasks, where the pertinent generalizations are only indirectly tied to the physical stimuli.

More generally, deeper links will be forged between the literature on implicit learning and language acquisition as the former begins to address the types of problems facing language learners (see, e.g., Perruchet & Vintner, 1998). One type of model which has been extensively applied to both implicit learning and language development is the simple recurrent network (SRN) (e.g., Christiansen, 1994; Cleeremans, 1993; Elman, 1990; Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1996). These models learn to predict the next element in a sequence, with additional information derived from the temporal context of preceding events via recurrent connections. Cleeremans and his colleagues (e.g., Cleeremans, 1993; Cleeremans & McClelland, 1991) demonstrated that the SRN model can account for human performance across a range of implicit sequence learning tasks. In particular, the SRN model exposed to finite-state grammars can acquire long-distance dependencies over embedded material, as long as subtle statistical properties of embedded strings depend on earlier information (D. Servan-Schreiber, Cleeremans, & McClelland, 1991). This type of task is highly relevant to language learning, and these results are mirrored by Elman's (1993) simulations, in which a similar model acquired a grammar containing recursively embedded relative clauses (under certain conditions discussed below).

To what extent might we expect the present findings to be captured by the SRN model, given that the acquisition of phrase structure has been cast here as a statistical learning problem? In particular, can these results be explained without implementing an additional constraint favoring predictive dependencies as cues to units? Because connectionist models are attuned to predictive information, it is possible that a SRN would learn in the same fashion as humans given the cues in this type of language (which could, in turn, offer a deeper computational explanation of the learning mechanisms described here). Alternatively, successful modeling of these data may require the implementation of additional constraints. Connectionist models acquiring natural-language-like structures often benefit from the consideration of constraints on human learners (e.g., Seidenberg & Elman, 1999); for example, the necessary conditions for successful acquisition of long-distance dependencies by Elman's (1993) model are derived from developmental constraints on working memory (Newport, 1990).

In conclusion, these results support the hypothesis that learners can detect predictive dependencies in the service of acquiring simple phrase structure, revealing a potential cue for the acquisition of phrase structure in natural language learning. Moreover, learners were not influenced by the variety of irrelevant surface patterns available in the input. Instead, learners may be constrained to detect a small subset of possible generalizations, thereby filtering out some of the many irrelevant generalizations available in the input. It is these constraints which turn the statistics of languages into a potential goldmine for learners, rather than a minefield of misleading information. To the extent that this type of view is correct, the striking similarities observed across human languages may reflect constraints on human learning abilities. The degree to which these constraints are tailored particularly for language learning, or instead emerge from other properties of cognition and perception, remains a key empirical question for future exploration.

APPENDIX 1: EXPOSURE AND TEST STIMULI FOR EXPERIMENT 1

Sentences are grouped according to their structures, rather than in the randomization orders presented during testing.

Exposure sentences

1. AE
   biff jux
   hep vot
   mib jux
   rud vot

2. ADE
   mib klor jux

3. ACF
   hep sig dupp
   mib neb loke

4. ADCF
   biff pell neb dupp
   rud klor cav loke
   biff klor lum loke
   rud pell sig dupp
   hep klor neb loke
   hep pell sig dupp
   mib pell cav loke

5. ACGF
   biff cav pilk dupp
   rud neb pilk loke

6. ADCGF
   hep klor lum tiz loke
   rud pell sig pilk dupp
   biff klor sig pilk loke
   mib pell neb tiz dupp

hep pell cav pilk loke
mib klor neb pilk dupp

7. AEC
rud jux lum
biff vot sig
mib jux sig

8. AECG
hep jux neb tiz
mib vot cav pilk
biff jux neb tiz

9. ADEC
rud pell vot lum

10. ADECG
mib pell vot neb tiz

11. ACFC
hep sig loke neb
mib neb dupp lum
hep sig loke cav
hep sig dupp sig

12. ACFCG
biff sig dupp neb tiz
rud lum loke cav pilk
biff cav dupp neb pilk
rud sig loke cav tiz

13. ADCFC
biff pell lum dupp cav
hep klor sig loke neb
hep pell lum loke cav
mib klor cav dupp neb
rud pell cav loke lum
mib klor neb loke sig

14. ACGFC
mib neb tiz loke sig
rud cav pilk dupp lum
biff neb pilk dupp cav
hep lum tiz loke cav
rud neb tiz dupp lum
hep vot lum

Rule Tests One and Two

Rule 1: Every sentence must contain at least one A word.

Rule Test One	*Rule Test Two*
mib sig dupp *sig dupp	hep cav loke *cav loke
biff lum loke sig tiz *lum loke sig tiz	biff lum dupp cav pilk *lum dupp cav pilk
hep jux lum *jux lum	mib vot neb *vot neb
rud lum tiz loke *lum tiz loke	rud sig tiz dupp *sig tiz dupp

Rule 2: No sentence may contain more than one A word.

Rule Test One	*Rule Test Two*
rud sig dupp *hep rud sig dupp	rud neb loke *hep rud neb loke
mib pell jux cav *mib biff pell jux cav	mib klor vot cav *mib biff klor vot cav
biff vot *biff rud vot	hep jux *hep rud jux
hep cav dupp sig *rud hep cav dupp sig	biff lum loke sig *rud biff lum loke sig

Rule 3: If there is an E word, there cannot be a CP.

Rule Test One	*Rule Test Two*:
hep jux *hep cav tiz jux	rud jux *rud cav pilk jux
biff vot neb *biff sig pilk vot neb	biff jux sig *biff lum tiz jux sig
rud klor jux *rud klor lum jux	hep pell vot *hep pell lum vot
mib pell vot sig *mib pell neb vot sig	mib klor vot cav *mib klor neb vot cav

Rule 4: If there is a D word, there must be an A word.

Rule Test One	*Rule Test Two*
hep klor jux cav *klor jux cav	mib klor vot cav *klor vot cav
biff pell lum loke *pell lum loke	biff pell neb loke *pell neb loke
rud pell neb dupp sig *pell neb dupp sig	hep klor sig dupp cav *klor sig dupp cav
mib klor vot neb *klor vot neb	rud pell jux neb *pell jux neb

Rule 5: If there is an F word, there must be a CP.

Rule Test One	*Rule Test Two*
mib pell sig loke *mib pell loke	hep cav pilk dupp *hep pilk dupp
rud pell lum loke cav *rud pell loke cav	biff sig tiz loke *biff tiz loke
biff neb dupp *biff dupp	rud jux neb tiz *rud jux tiz
hep cav dupp lum *hep dupp lum	mib neb pilk dupp lum *mib pilk dupp lum

Rule 6: If there is a G word, there must be a C word.

Rule Test One	*Rule Test Two*
biff cav tiz dupp *biff tiz dupp	rud pell cav loke *rud pell loke
hep sig pilk loke *hep pilk loke	mib klor lum dupp sig *mib klor dupp sig
mib vot cav tiz *mib vot tiz	biff lum loke *biff loke
rud neb pilk loke cav *rud pilk loke cav	hep cav loke neb *hep loke neb

Fragment Test:

Asterisks signal fragments spanning a phrase boundary, not ungrammaticality.

Fragments testing the A phrase:

*AD versus *AC*	*AD versus *DC*
biff klor *biff cav	hep klor *klor sig

rud pell *rud neb	hep pell *pell lum
mib pell *mib neb	rud klor *klor cav
hep klor *hep sig	biff pell *pell neb

Fragments testing the B phrase:

*CGF versus *GFC*	*CGF versus *DCG*
sig pilk loke *pilk loke neb	neb tiz dupp *klor neb tiz
lum tiz dupp *tiz dupp cav	cav pilk loke *pell cav pilk
cav pilk dupp *pilk dupp lum	neb pilk dupp *klor neb pilk
lum tiz loke *tiz loke neb	lum tiz loke *klor lum tiz

Fragments testing the C phrase:

*CG versus *DC*	*CG versus *FC*
lum tiz *klor lum	sig tiz *loke sig
neb tiz *pell neb	neb tiz *dupp neb
sig pilk *klor sig	cav pilk *loke cav
neb pilk *klor neb	sig pilk *dupp sig

REFERENCES

Altmann, G. T. M., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21,** 899–912.

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.

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* (pp. 117–134), Hillsdale, NJ: Erlbaum.

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

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

Braine, M. D. S. (1971). On two types of models of the internalization of grammars. In D. I. Slobin (Ed.), *The ontogenesis of grammar: A theoretical symposium* (pp. 153–186), New York: Academic Press.

Braine, M. D. S., Brody, R. E., Brooks, P. J., Sudhalter, V., Ross, J. A., Catalano, L., & Fisch, S. (1990). Exploring language acquisition in children with a miniature artificial language: Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of Memory and Language*, **29,** 591–610.

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.

Canfield, R. L., & Haith, M. M. (1991). Young infants' visual expectations for symmetric and asymmetric stimulus sequences. *Developmental Psychology*, **27,** 198–208.

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.

Cleeremans, A. (1993). *Mechanisms of implicit learning: Connectionist models of sequence processing*. Cambridge, MA: MIT Press.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, **120,** 235–253.

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

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

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Fathman, A. (1975). The relationship between age and second language productive ability. *Language Learning*, **25,** 245–253.

Gómez, R. L. (1997). Transfer and complexity in artificial grammar learning. *Cognitive Psychology*, **33,** 154–207.

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

Gómez, R. L., Gerken, L. A., & Schvaneveldt, R. W. (2000). The basis of transfer in artificial grammar learning. *Memory & Cognition*, **28,** 253–263.

Gómez, R. L., & Schvaneveldt, R. W. (1994). What is learned from artificial grammars? Transfer tests of simple associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20,** 396–410.

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.

Hasher, L. & Zacks, R. T. (1984). Automatic processing of fundamental information. *American Psychologist*, **39,** 1372–1388.

Hasher, L., Zacks., R. T., Rose, K. C., & Sanft, H. (1987). Truly incidental encoding of frequency information. *American Journal of Psychology*, **100,** 69–91.

Jackendoff, R. S. (1977). *X-bar syntax: A study of phrase structure*. Cambridge, MA: MIT Press.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language acquisition: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, **21,** 60–99.

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., & Martin, S. (1994). Domain-general abilities applied to domain-specific tasks: Sensitivity to probabilities in perception, cognition, and language. *Lingua*, **92,** 105–140.

Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **22,** 169–181.

Krashen, S., Scarcella, R., & Long, M. (1982). *Child-adult differences in second language acquisition*. Rowley, MA: Newbury House.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior: The Hixon Symposium* (pp. 112–146), 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, pp. 127–151), New York: Gardner Press.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven month-old infants. *Science*, **283,** 77–80.

McAndrews, M. P., & Moscovitch, M. (1985). Rule-based and exemplar-based classification in artificial grammar learning. *Memory & Cognition*, **19,** 469–475.

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.

Moeser, S. D., & Bregman, A. S. (1972). The role of reference in the acquisition of a miniature artificial language. *Journal of Verbal Learning and Verbal Behavior*, **11,** 759–769.

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.

Muelemans, T., & Van der Linden, M. (1997). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23,** 1007–1028.

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

Perruchet, P. (1994). Defining the knowledge units of a synthetic language: Comment on Vokey and Brooks (1992). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20,** 223–228.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General*, **119,** 264–275.

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.

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Verbal Learning and Verbal Behavior*, **81,** 317–327.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York: Oxford University Press.

Reber, A. S., & Allen, R. (1978). Analogic and abstraction strategies in synthetic grammar learning. *Cognition*, **6,** 189–221.

Reber, A. S., & Lewis, S. (1977). Implicit learning: An analysis of the form and structure of a body of tacit knowledge. *Cognition*, **5,** 331–361.

Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. *Journal of Experimental Psychology: General*, **125,** 123–138.

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

Rovee-Collier, C. (1991). The "memory system" of prelinguistic infants. In The Development and Neural Bases of Higher Cognitive Functions, A. Diamond, (Ed.), *Annals of the New York Academy of Sciences*, **608,** 517–536.

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

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.

Seidenberg, M. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science*, **275,** 1599–1603.

Seidenberg, M., & Elman, J. (1999). Do infants learn grammar with algebra or statistics? *Science*, **284,** 433.

Servan-Schreiber, D., Cleeremans, A., & McClelland, J. L. (1991). Graded state machines: The representation of temporal contingencies in simple recurrent networks. *Machine Learning*, **7,** 161–193.

Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **16,** 592–608.

Shanks, D. R., Johnstone, T., & Staggs, L. (1997). Abstraction processes in artificial grammar learning. *Quarterly Journal of Experimental Psychology*, **50,** 216–252.

Slavoff, G. R., & Johnson, J. S. (1995). The effects of age on the rate of learning a second language. *Studies in Second Language Acquisition*, **17,** 1–16.

Tunney, R. J., & Altmann, G. T. M. (1999). The transfer effect in artificial grammar learning: Reappraising the evidence on the transfer of sequential dependencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **25,** 1322–1333.

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

Vokey, J. R., & Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammar. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **18,** 328–344.

Whittlesea, B. W. A., & Dorken, M. D. (1993). Incidentally, things in general are particularly determined: An episodic-processing account of implicit learning. *Journal of Experimental Psychology: General*, **112,** 227–248.

Younger, B. A. (1985). The segregation of items into categories by ten-month-old infants. *Child Development*, **54,** 858–867.

(Received January 11, 2000)

(Revision Received August 25, 2000; published online March 15, 2001)