

Constraints on Statistical Language Learning

Jenny R. Saffran

University of Wisconsin—Madison

How do learners discover the structure in linguistic input? One set of cues which learners might use to acquire phrase structure are the dependencies, or predictive relationships, which link elements within phrases. In order to determine whether learners can use this statistical information, adults and children were exposed to artificial languages that either contained or violated the kinds of dependencies that characterize natural languages. Additional experiments contrasted the acquisition of these linguistic systems with the same grammars implemented as non-linguistic input (sequences of nonlinguistic sounds or shapes). Predictive relationships yielded better learning for sequentially presented auditory stimuli, and for simultaneously presented visual stimuli, but no such advantage was found for sequentially presented visual stimuli. Learning outcomes were not affected by the degree to which the input contained linguistic content. These findings suggest that constraints on learning mechanisms that mirror the structure of natural languages are not tailored solely for language learning. Implications for theories of language acquisition and perceptual learning are discussed. © 2002 Elsevier Science (USA)

Key Words: statistical learning, language, syntax, phrase structure.

One of the central questions in the field of language acquisition is the nature of the mechanisms that underlie the transfer of information from the child's linguistic environment to the child's mind. The range of mechanisms proposed to subserve this process mirrors the complexity of the knowledge that children eventually possess about their native language. In the present research, we focused on one type of mechanism hypothesized to underlie aspects of language acquisition: the process of statistical learning, or the detection of patterns of sounds,

words, and classes of words in the service of discovering underlying structure.

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 reemerged 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, the widespread availability of large corpora of child-directed speech, and empirical research demonstrating that humans can perform statistical language learning tasks in laboratory experiments (e.g., Cartwright & Brent, 1997; Elman, Bates, Karmiloff-Smith, Parisi, & Plunkett, 1996; Gómez & Gerken, 1999; Goodsitt, Morgan, & Kuhl, 1993; MacWhinney, 1999; Mintz, Newport, & Bever, 1995; Redington, Chater, & Finch, 1998; Saffran, 2001; Saffran, Aslin, & Newport, 1996a; Saffran, Newport, & Aslin, 1996b; Seidenberg, 1997; Seidenberg & MacDonald, 1999). Indeed, the emerging body of evidence suggests that humans, including infants, may be exceptionally skilled statistical learners.

The capacity to detect the statistical properties of linguistic input is likely to be a useful component of the language learner's arsenal of

Experiment 1 was submitted to the University of Rochester as a portion of the author's dissertation and was supported by NIH Training Grant 5T32DC0003 to the University of Rochester and by NIH Grant DC00167 to Elissa Newport. Experiments 2–6 and the preparation of the manuscript were supported by grants to the author from NIH (HD03352 and HD37466), NSF (BCS-9983630), and the University of Wisconsin-Madison Graduate School. I am grateful to Maria Boardman, Toby Calandra, Carrie Franz, Elizabeth Johnson, and Lana Nenide and the members of the UW-Madison Learning Languages Lab for their assistance in conducting these experiments, to three anonymous reviewers for helpful suggestions on a previous draft, and to Martha Alibali, Dick Aslin, Art Glenberg, Michael Kaschak, Lana Nenida, Erik Thissen, and especially Elissa Newport for helpful discussions.

Address correspondence and reprint requests to Jenny R. Saffran at the Department of Psychology, University of Wisconsin-Madison, 1202 West Johnson Street, Madison, WI 53706. E-mail: jsaffran@facstaff.wisc.edu.

acquisition devices. However, for statistical learning to be a viable component of language acquisition, learners must be able to detect input statistics that are pertinent to linguistic structure amid all the irrelevant information in the input. To do so, statistical learning mechanisms must be *constrained or biased* to preferentially perform certain kinds of computations over certain kinds of input. The pertinent generalizations to be drawn from a linguistic corpus are surrounded by irrelevant possible generalizations. Any learning device without the right architectural, representational, or computational constraints risks being sidetracked by the 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, in principle: 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, such that learners perform only a subset of the logically possible computations. Learning in biological systems is limited by internal factors; there are species differences in the specific types of stimuli that serve as privileged input (e.g., Garcia & Koelling, 1966; Marler, 1991). External factors also strongly bias learning, because input from structured domains consists of nonrandom information. In order for statistical learning accounts to succeed, language 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 acquisition or from more general cognitive and/or perceptual constraints on human learning.

A related issue pertaining to learning-based theories of language acquisition concerns the nature of language itself. Human languages are

remarkably similar to one another, despite surface differences. How does this long-standing observation mesh with the hypothesis that learning plays a central role in language acquisition? In particular, an overly powerful learning device should readily learn structures that are *not* present in natural languages, as well as those structures that are ubiquitous in human languages. The solution explored here is that the learning mechanisms applied to language may be constrained to preferentially learn certain types of patterns (e.g., Bever, 1970; Christiansen, 1994; Christiansen & Devlin, 1997; Ellefson & Christiansen, 2000; Morgan, Meier, & Newport, 1987; Newport, 1982, 1990). If the structures that are most learnable are also those that recur cross-linguistically, then the similarity of human languages may have roots in the learning process itself: constraints on language learning may shape the structure of natural languages.

To explore these issues, the current experiments address the hypothesis that statistical learning is constrained: learners are most likely to track those statistical properties of language that will afford the discovery of natural language structure. The aspect of language addressed by these studies is 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 grouping, (*The professor*) (*graded (the exam)*), reflects English phrase structure, which generates a nonlinear hierarchically organized structure. Hierarchical phrase structure represents a fascinating learning problem, because the child must somehow arrive at nonlinear structure that 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 and other types of grouping cues may also serve to chunk the input into phrasal units (e.g., Morgan, Meier, & Newport, 1987, 1989).

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, 2001).

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 detect the types of structures observed in natural languages. In the current research, we focused on the possibility that learners may preferentially acquire the predictive dependencies consistently observed in natural languages. Predictive dependencies may be recast as conditional probabilities, a type of statistic known to be pertinent to learners across domains (e.g., Aslin, Saffran, & Newport, 1998; Rescorla, 1966). To the extent that predictive dependencies and human learning mechanisms are a good fit, we would expect that learners exposed to languages containing predictive dependencies (like natural languages) would outperform learners exposed to languages that lack predictive dependencies (unlike natural languages). Learners confronted with serially presented input may be constrained to detect the predictive relations between different lexical categories (amid all the other statistical properties of the sequence of words and word categories), which in turn would facilitate the detection of phrase structure. If this is the case, then one reason languages may contain predictive depend-

encies, along with other types of cues to phrase structure, is that they enhance learnability (e.g., Morgan et al., 1987). We can then ask whether the use of predictive dependencies is a constraint on *language* learning or whether this mechanism also operates over material from other domains.

To address these questions, we contrasted the acquisition of two artificial languages in a series of six experiments. One of the languages contained predictive dependencies, while the other did not; both languages also contained many other statistical properties. After exposure, we assessed learning outcomes for the two languages. We began by testing adult learners in Experiment 1. Experiment 2 extended the investigation of the role of predictive dependencies in language learning to include child learners. In Experiments 3–5, we assessed the domain-generality of the hypothesized constraint on learning using materials drawn from nonlinguistic domains. Experiment 6 further explored modality differences by examining the effects of simultaneous versus sequential presentation on detecting predictive dependencies in visual tasks. The overarching goals of these investigations were to ask whether predictive dependencies affect the learnability of sequential structure and to assess the domain-generality of this constraint on learning.

EXPERIMENT 1

To investigate the contributions of predictive dependencies to language acquisition, we contrasted the acquisition of two artificial grammars. One of these grammars, Language P, contained predictive dependencies as cues to phrase structure; if one member of a phrase was present, the other member was always present. The other grammar, Language N, did not contain predictive dependencies as a cue to phrase structure; the presence of one member of a phrase did not predict the presence of the other member. Following exposure, we assessed language learning using the same test for all participants. If predictive dependencies assist in learning basic syntax, then participants acquiring Language P should outperform participants acquiring Language N.

Method

Participants

Forty monolingual English-speaking undergraduates at the University of Rochester participated in this study. Subjects were randomly assigned to the two experimental conditions. Three additional subjects (one from the Language P condition and two from the Language N condition) were excluded from the analysis for making errors on the practice trials presented immediately prior to testing. All subjects in this and all subsequent experiments gave informed consent.

Description of the Linguistic Systems

The artificial grammars were adapted from the languages used by Morgan and Newport (1981) and Saffran (2001); exposure sentences are listed in Appendix 1. Each letter in the grammar represents one form class, consisting of two to four monosyllabic nonsense words (see Table 1).

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 (e.g., if D is present, A must be present).

(1) Language P

$$S \rightarrow AP + BP + (CP)$$

$$AP \rightarrow A + (D)$$

$$BP \rightarrow CP + F$$

$$CP \rightarrow C + (G)$$

Language P contains the type of predictive structure found in natural languages. In A phrases, A words can occur without D words, but 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; the optional CP was necessary to balance the languages in terms of sentence types), but if an F word is present, a C phrase must precede it. The conditional probability of A|D is 1.0; the same is true of the other within-phrase pairs in the language.

Importantly, the directionality of the statistical patterns in Language P is the opposite of the native language of our participants. In English, 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). Language P employed the opposite pattern: the predictor D follows A, G follows C, and F follows the C phrase. Any attempt to project English structure onto the artificial language should have resulted in poor learning outcomes.

The second language did not contain predictive cues to phrase boundaries (Language N, for nonpredictive). This grammar was characterized by overarching optionality: the presence of one word type never predicted the presence of another. Note, however, that Language N still possesses phrase structure of a sort—the *absence* of one word type within a phrasal unit predicts the presence of another (e.g., if A is *not* present, D must be present). Language N contained the same form classes and vocabulary as Language P (see Table 1).

(2) Language N:

$$S \rightarrow AP + BP$$

$$AP \rightarrow [(A) + (D)] \text{ (must have at least one; if both, A precedes D)}$$

$$BP \rightarrow CP + F$$

$$CP \rightarrow [(C) + (G)] \text{ (must have at least one; if both, C precedes G)}$$

Languages P and N are similar on other pertinent dimensions. Both languages contained the same number of grammatical categories and vocabulary items. Language N generates fewer sentence types (9) than Language P (12). For the purpose of these experiments, only sentence types with five or fewer words were used (eight types for Language P, nine for Language N). Language N also had shorter sentences on average: Language P generated 60% more five-word sentences than Language N and only 40% as many three-word sentences. Importantly, the two-word pairs (phrases) that were manipulated during testing (AD and CG) occurred equally often in both languages.

TABLE 1

Word Categories from the Artificial Language

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

A trained female speaker produced 50 sentences from each language, chosen randomly with the constraint that each word occurred with similar frequency in both languages and that AD and CG occurred equally often in both languages. Each sentence list was recorded in two random orders, with uniformly descending prosody across each sentence. Words occurred at a rate of approximately two words per second. Approximately 2s of silence separated each sentence. The speech was recorded using a Sony Walkman Pro tape deck. Each recorded block consisted of 100 sentences (the two orderings of the 50 sentences) and was approximately 7 min in duration.

Procedure

Participants were exposed to either Language P or Language N in an incidental learning paradigm used previously by Saffran et al. (1997) and Saffran (2001), to minimize the effects of strategic learning processes. While participants listened to the exposure materials (via a Sony tape deck and speakers), they were asked to create an illustration using the children's computer coloring game KidPix2 on a Mac Quadra. Participants were informed that there would be a nonsense language playing in the background, but were not informed about the structure of the language. We also informed participants that they would be tested on the nonsense language, but did not tell them which aspects of the language would be tested. Because participants knew they would be tested, this procedure was not fully incidental. All participants were tested individually in a single session, hearing the 7-min recorded block of 100 sentences (from either Language

P or Language N) four times, with a short break after the second repetition. After the fourth and final repetition of the sentences, subjects received a test designed to examine their learning of the rules.

Rule test. In order to test the effects of predictive dependencies on language learning, participants exposed to Languages P and N received the same test. Each test item included a pair of sentences: a novel grammatical sentence and an ungrammatical sentence, recorded by the speaker who recorded the exposure materials. To contrast the two groups of language learners, the grammatical items were legal in both languages, and the ungrammatical items were illegal in both languages (see Table 2).¹ The test items thus assessed the acquisition of rules common to both languages. Test sentences are listed in Appendix 2. After hearing each sentence pair, participants were asked to determine whether the first or the second sentence in the pair sounded more like the exposure language and to mark their response on an answer sheet.

Each of the four rules was tested by six novel sentence pairs, rendering 24 forced-choice trials. Participants received four practice trials preceding the test in order to clarify the test instructions: two trials in English and two in the nonsense language (with incorrect sentences consisting of scrambled word order). These practice trials also provided exclusion criteria to ensure that learners were attending during exposure; participants who made errors were excluded from the analysis.

Results and Discussion

The first analysis asked whether subjects succeeded in learning Language P and Language N. Each group's overall performance was significantly better than would be expected by chance: for Language P, the total score was 17.9 of a possible 24: $t(19) = 10.42$, $p < .0001$; for Language N, the total score was 16.2: $t(19) = 7.44$,

¹An additional rule was included in the test administered in all five experiments. However, because this rule only applied to the structure of Language P, the results from items testing this rule were not included in the analyses reported in this paper. Inclusion of this rule does not change the overall pattern of results.

TABLE 2
Rules Tested in Experiments 1–5

Rule	
1	Sentences must contain an A phrase. BIFF KLOR SIG PILK JUX [A-D-C-G-F] *SIG PILK JUX [C-G-F]
2	D words follow A words, while G words follow C words. HEP PELL LUM PILK JUX [A-D-C-G-F] *HEP PILK LUM PELL JUX [A-G-C-D-F]
3	Sentences must contain an F word. MIB LUM PILK VOT [A-C-G-F] *MIB LUM PILK [A-C-G]
4	C phrases must precede F words. RUD PELL NEB DUPP [A-D-C-F] *RUD PEL DUPP [A-D-F]

$p < .0001$. Table 3 presents subjects' mean scores on the individual rules tested.

Our main hypothesis concerned differences in learning as a function of structural differences between the two languages. To address this question, the overall scores for the two language groups were compared using an ANOVA. Language P learners outperformed Language N

learners: $F(1,38) = 4.52, p < .05$. This difference suggests that Language P was easier for subjects to acquire than Language N.

Because all learners received the same test, it is unlikely that features of the test itself differentially influenced Language P and N learners. However, there remains the possibility that surface variables in the exposure sentences influ-

TABLE 3
Mean Scores and Significance Tests (Two-Tailed) against Chance (Three of Six Possible), for Language P and Language N, Experiments 1–6

Experiment No.		Rule			
		1	2	3	4
Language P					
1	Linguistic auditory (adult)	4.40 **	4.00**	4.75**	4.75**
2	Linguistic auditory (child)	4.60**	3.20	5.20**	4.27**
3	Nonlinguistic auditory	4.52**	4.06**	3.98**	4.59**
4	Linguistic visual	4.45**	3.55**	4.53**	4.53**
4	Nonlinguistic visual	5.19**	4.19**	4.31**	4.27**
5	Nonlinguistic auditory	4.74**	3.78**	3.82**	4.67**
5	Nonlinguistic visual	4.58**	3.92**	4.21**	4.04**
6	Simultaneous visual	5.43**	5.68**	5.00**	4.96**
Language N					
1	Linguistic auditory (adult)	3.25	3.60*	5.35**	4.00**
2	Linguistic auditory (child)	3.07	3.33	3.80*	3.80*
3	Nonlinguistic auditory	3.71**	3.97**	4.54**	3.66**
4	Linguistic visual	3.90**	3.40	4.70**	3.93*
4	Nonlinguistic visual	4.72**	4.08**	4.52**	3.84**
5	Nonlinguistic auditory	3.64*	3.28	3.52	3.96*
5	Nonlinguistic visual	4.75**	3.63**	4.67**	3.72**
6	Simultaneous visual	5.18**	5.89**	4.15**	3.96**

* $p < .05$.

** $p < .01$.

enced performance during testing. In prior research using a very similar grammar (Saffran, 2001), we used analyses of covariance to rule out a number of surface variables which might have influenced mapping between exposure and test items, including bigram frequencies or chunk strength (e.g., Knowlton & Squire, 1996; Perruchet & Pacteau, 1990; Redington & Chater, 1996; Servan-Schreiber & Anderson, 1990), frequencies of beginning and ending bigrams, or anchor strength (e.g., Perruchet, 1994; Reber & Lewis, 1977), legality of the first element (e.g., Reber & Allen, 1978; Tunney & Altman, 1999), presence of unique chunks (e.g., Meulemans & Van der Leden, 1997), and overall similarity to individual exposure strings (e.g., McAndrews & Moscovitch, 1985, Vokey & Brooks, 1992). Because some of these factors may have differed in the exposure sentences for Languages P and N, we entered the current data into an analyses of covariance (ANCOVA) in which string and substring features were entered as covariates. The question of interest was whether grammatically (whether or not a given test item violated a rule of the language) would continue to exert differential effects on the two language groups' (P versus N) performance, as measured by a significant grammaticality by language interaction, once other factors representing surface characteristics of the stimuli were entered into the model.

The test consisted of 24 forced-choice pairs contrasting grammatical and ungrammatical items, rendering 48 items for the ANCOVA from each language condition. Language P and Language N scores for each item were included separately, rendering a total of 96 items for the 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. Grammaticality was coded as a two-level factor: items were either grammatical or not. Language (P versus N) and legality of the first word were also coded as two-level factors. The remaining factors were all continuous variables computed for each test item relative to the exposure corpus from either Language P or N: chunk strength (the average of

the input frequencies for all word pairs for each item), anchor strength (the composite of the input frequencies for the initial and final word pairs for each item), uniqueness (the 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. As noted previously, Language P sentences were longer, on average, than Language N sentences (P, 4.24 words; N, 3.88 words). On the test, grammatical sentences were longer, on average, than ungrammatical sentences (grammatical, 4.71 words; ungrammatical, 3.58 words). The Language P sentences were thus closest in length to the grammatical sentences, whereas the Language N sentences were closest in length to the ungrammatical sentences. This imbalance raises the possibility that Language P learners outperformed Language P learners outperformed Language N learners because their input was closest in length to the grammatical sentences, while Language N input was closest in length to the ungrammatical sentences.

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. 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 that included a covariate were eliminated from the final models.

The final model thus consisted of three factors and six covariates, and the interaction term for the two main effects (Grammaticality \times Language). As shown in Table 4, the main effects of Grammaticality [$F(1,90) = 74.4$] and First Word Legality [$F(1,90) = 5.58$] were significant, as was the interaction between Grammaticality and Language [$F(1,90) = 7.22$]. These results suggest that other than the legality of the first word, surface variables did not contribute to subjects' endorsement of items, and grammaticality continued to exert effects even when the variance accounted for by the surface

TABLE 4
ANCOVA *F* Values for Experiments 1–5

Factor	Experiment No.						
	1 [Linguistic auditory] (adult)	2 [Linguistic auditory] (child)	3 (Nonlinguistic auditory)	4 (Linguistic visual)	4 (Nonlinguistic visual)	5 (Nonlinguistic auditory 2)	5 (Nonlinguistic visual 2)
Grammaticality	74.4**	62.9**	74.5**	107.4**	118.7**	26.4*	44.8**
Language	0.08	0.29	0.11	0.96	0.03	0.07	0.03
Grammaticality × language	7.2**	35.5**	9.99**	16.75**	0.01	9.35**	0.59
Length	0.01	0.27	0.15	0.02	0.89	0.01	2.14
First word legal	5.58*	2.02	0.15	0.96	5.09*	0.01	2.46
Chunk strength	1.64	0.30	0.89	3.31	0.55	0.74	1.03
Anchor strength	.34	0.09	3.26	0.79	0.45	2.74	1.20
Uniqueness	0.30	7.66**	0.01	0.44	0.01	0.25	0.69
Similarity	1.07	2.89	0.36	5.25*	0.07	0.23	2.02

**p* < .05.

***p* < .01.

df = 1, 90.

variables was removed. More importantly, the significant interaction between Grammaticality and Language indicates that Language P and N learners showed different levels of responses to items as a function of their grammaticality. If length or other surface variables differentially affected the two conditions, then we would have expected the Grammaticality × Language interaction to be removed when these variables were included in the analysis. Instead, the results suggest that the surface variables cannot explain the differential performance of Language P versus Language N learners.

The results of Experiment 1 suggest that the availability of predictive dependencies in the input assists rudimentary language learning—or, conversely, that a lack of predictive dependencies impedes learning. Clearly, it is not the case that languages lacking predictive dependencies are unlearnable; participants acquiring Language N exceeded chance performance. However, the lack of predictive dependencies impaired overall learnability relative to Language P, at least given the exposure and test items used in this experiment. These findings suggest that learners may take advantage of the dependencies that characterize natural language phrase structure in the course of language acquisition.

An immediate question raised by these findings is whether adult strategic learning processes led to the P versus N performance difference. Despite the use of the incidental procedure, it is possible that our adult participants noticed the optional elements in the Language N input and were misled to believe that they were being exposed to random structures, rendering poorer outcomes. A related question concerns the hypothesized constraint to detect and use predictive dependencies. In order for this bias to assist learners acquiring their native language, it must be present during childhood. To address these two issues, the next experiment compared child learners exposed to Language P and Language N.

EXPERIMENT 2

Method

Children are less likely than adults to impose learning strategies in artificial grammar learning

tasks or to attempt to “translate” the nonsense input into their native language. Thus, results from child learners are unlikely to reflect lab-induced learning strategies. Prior research showing the difficulty of eliciting metalinguistic judgments from young children (e.g., Slavoff & Johnson, 1995) led us to test children who were older than 7 years 6 months, but still within the critical period for language learning (had we used older children, it would be unclear whether the results of our experiments are pertinent to child language learners). Based on prior research using similar procedures with children (Saffran, 2001), we anticipated that the adults would outperform the children due to the task demands induced by the forced-choice testing procedure. However, we hypothesized that children, like adults, would show enhanced test performance when predictive cues to phrase structure were available during learning over those when they were not.

Participants

Thirty monolingual English-speaking children were recruited from after-school programs in Madison, Wisconsin. The children ranged in age from 7 years 6 months to 9 years 8 months and were randomly assigned to the two experimental conditions (Language P mean age, 8 years 3 months; Language N mean age, 8 years 1 month). Parents gave informed consent prior to testing.

Procedure

The children were exposed to either Language P or Language N from Experiment 1. As in Experiment 1, we used an incidental learning paradigm. However, because results from prior studies on syntax learning suggested that the cover task of coloring on the computer might be too engaging for the children (Saffran, 2001), we gave the children quiet toys to play with during exposure (Legos, Etch-a-Sketch, and coloring books). As in Experiment 1, we told the children that there would be a nonsense language playing in the background and that they would be tested later in the study, but they were told nothing about the structure of the language. Exposure was otherwise identical to that in Experiment 1. Testing was identical to that in Experi-

ment 1, except that the children received as many practice trials in English as necessary to ensure that they understood the procedure, as well as additional practice trials using the nonsense words; the children received stickers as a motivator after every third test trial.

Results and Discussion

The first analysis asked whether the children succeeded in learning Language P and Language N. Each group’s overall performance was significantly better than would be expected by chance: for Language P, the total score was 17.27 of a possible 24, $t(14) = 6.30$, $p < .0001$; for Language N, the total score was 14, $t(14) = 2.24$, $p < .05$. Table 3 presents subjects’ mean scores on the individual rules tested.

To contrast performance on Language P versus Language N, the overall scores for the two language groups were contrasted in an ANOVA. Language P learners outperformed Language N learners: $F(1,28) = 7.12$, $p < .05$. This difference suggests that Language P was easier for children to acquire than Language N. As in Experiment 1, we submitted the results to an ANCOVA to determine whether surface variables could account for the P versus N difference. As shown in Table 4, Grammaticality [$F(1,90) = 62.9$] and Uniqueness [$F(1,90) = 7.66$] were significant, as was the interaction between Grammaticality and Language [$F(1,90) = 32.5$]. Like the adults, the children’s differential performance on Languages P and N was not a function of surface features of the exposure and test items.

We next compared the children’s performance with that of the adults from Experiment 1. A two-factor ANOVA including age (child versus adult) and language (P versus N), with total score as the dependent measure, revealed main effects of Age [$F(1,66) = 4.07$, $p < .05$] and Language [$F(1,66) = 12.51$, $p < .001$], with a nonsignificant interaction between Age and Language [$F(1,66) = 1.24$, n.s.]. While adults performed better than children overall, the effects of predictive dependencies emerged in both age groups. Although these children are beyond the age at which first language syntax is typically acquired, the results suggest that predictive dependencies may be available for use in

the process of first language acquisition. Future work will extend investigations of predictive dependencies to include late infancy and toddlerhood; it has recently been demonstrated that infants as young as 12 months (Gómez & Gerken, 1999) and even 8 months (Marcus, Vijayan, Bandi Rao, & Vishton, 1999) can learn rudimentary syntactic patterns generated by artificial grammars.

The findings from Experiments 1 and 2 support the hypothesis that predictive dependencies play a role in acquiring rudimentary syntax in language learning. We can then ask whether detecting structures using dependencies between classes of items is a learning process particularly tailored for linguistic input or whether this learning mechanism can operate over materials drawn from other domains. 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 a more general feature of the acquisition of serially presented information. To directly test the domain specificity of this learning process, we contrasted the acquisition of Languages P and N using nonlinguistic materials. Participants received auditory exposure to “languages” in which the “words” were distinctive nonlinguistic computerized sounds. We then asked whether Language P learners would continue to outperform Language N learners given nonlinguistic input.

EXPERIMENT 3

To assess the role of predictive cues in nonlinguistic auditory learning, we translated Languages P and N into a vocabulary of nonlinguistic sounds. All other aspects of the experiment were identical to those of Experiment 1. We hypothesized that if predictive cues afford learnability benefits for structures other than language, then Language P learners should outperform participants exposed to Language N.

Method

Participants

Eighty-one monolingual English speaking undergraduates at the University of Wisconsin-

Madison participated in this study participated in this study for course extra credit. Forty-six of the participants were assigned to Language P and 35 were assigned to Language N.²

Materials

To create nonlinguistic auditory stimuli, we translated Languages P and N into nonlinguistic sounds drawn from the digitized bank of alert sounds provided with Windows 98. Each word corresponded to a different sound, chosen to be highly 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 stimuli in Experiment 1. The stimuli were combined for presentation using SuperLab software running on a PowerPC. “Words” occurred at a rate of approximately two per second, with two sec of silence separating each sentence. The stimuli were recorded from the computer using a Sony Minidisk recorder for playback to experimental participants. Following exposure, participants received the forced-choice test used in Experiment 1, translated into nonlinguistic sounds. No linguistic information was available for learners during exposure or testing.

Procedure

The procedure was identical to that of Experiment 1, except that the cover task of coloring on the computer during exposure was not used; we planned to contrast the auditory nonlinguistic materials from Experiment 3 with visual nonlinguistic materials (Experiment 4), and we could not use a visual cover task with the visual learning tasks.

²The imbalance in subject assignments reflects the prior use of the exclusion criterion described in Experiment 1: participants making errors on the practice test were originally excluded from the analyses. As a very large number of participants (37) were excluded by this criterion, it is likely that the criterion was overly conservative; the results reported in Experiments 3–6 include all participants tested regardless of their performance on the practice test. In all cases, the results of the analyses are unaffected by the inclusion of the previously excluded participants.

Results and Discussion

The first analysis asked whether subjects succeeded in learning Language P and Language N. Each group's overall performance was significantly better than would be expected by chance: for Language P, the total score was 17.52 of a possible 24, $t(45) = 11.87$, $p < .0001$; for Language N, the total score was 15.88, $t(34) = 8.73$, $p < .0001$. Table 3 presents subjects' mean scores on the individual rules tested.

To assess differences in learning as a function of structural differences between the two languages, we contrasted the overall scores for the two language groups in an ANOVA. Language P learners significantly outperformed Language N learners, $F(1,79) = 4.03$, $p < .05$. As in the linguistic task used in Experiments 1 and 2, Language P was easier for subjects to acquire than Language N. To ensure that this pattern of results was not due to surface variables, we applied the ANCOVA model from Experiment 1 to these data. As shown in Table 4, the only significant effects were Grammaticality [$F(1,90) = 74.45$] and the Grammaticality \times Language interaction [$F(1,90) = 9.99$], supporting the hypothesis that the differential performance of the Language P and N groups was due to structural properties of the two languages.

To determine whether linguistic and nonlinguistic auditory materials are learned differently, we contrasted the results from the present experiment with the findings from Experiment 1. The only difference between the two experiments lies in their vocabularies, which were nonsense words in Experiment 1 and nonsense sounds in Experiment 3. The grammars (Languages P and N) and test materials were identical. We submitted the total scores from Experiments 1 and 3 to a two-factor ANOVA including domain (linguistic versus nonlinguistic) and language (P versus N). The analysis revealed a main effect of Language [$F(1,117) = 7.88$, $p < .01$], with Language P learners outperforming Language N learners. The main effect of Domain [$F(1,80) = 1.01$, n.s.] and the interaction between Domain and Language [$F(1,80) = .16$, n.s.] were not significant. Thus, Language P was learned better than Language

N regardless of the linguistic status of the materials.

The results of the first three experiments suggest that predictive dependencies support learning, even when the input is nonlinguistic. These findings mirror prior results suggesting that the computation of sequential transitional probabilities in word segmentation tasks can occur whether "words" are created from syllables (Saffran, Newport, & Aslin, 1996b; Saffran, Aslin, & Newport, 1999) or nonlinguistic tones (Saffran, Johnson, Aslin, & Newport, 1999; Saffran & Griepentrog, 2001). The ability to use sequential transitional probabilities in word segmentation has also been demonstrated across modalities: learners can track the transitional probabilities between elements when presented with visuomotor patterns (Hunt & Aslin, 2001) and visuospatial patterns (Fiser & Aslin, 2001).

In Experiment 4, we asked whether the availability of predictive dependencies would affect learning across modalities. If the constraint to detect predictive dependencies is domain-general, then the modality within which the input is implemented should not affect learning, and materials containing predictive cues to phrase structure should be learned better than materials that do not. We thus anticipated that learners exposed to Language P presented visually would outperform learners acquiring Language N.

EXPERIMENT 4

This study is a conceptual replication of Experiments 1 and 3 in the visual domain. Learners were presented with either visual nonsense words or visual nonsense shapes, following the grammars of either Language P or Language N. The timing parameters for the sequential presentation of visual forms were identical to those used for the presentation of auditory forms in Experiments 1–3. Following exposure, learners received the test used in the previous experiments, implemented in either visual nonsense words or visual nonsense shapes. If predictive dependencies assist learners in acquiring basic syntactic structure in visual learning tasks, then participants acquiring

Language P should outperform participants acquiring Language N. We can also assess the effects of linguistic versus nonlinguistic stimuli in the visual domain.

Method

Participants

One-hundred and seven monolingual English speaking undergraduates at the University of Wisconsin-Madison participated in this study for course extra credit. Fifty-six participants were assigned to the Linguistic Visual condition, and 51 were assigned to the Nonlinguistic Visual condition. Within each condition, the participants were assigned to either Language P or Language N.

Materials

To create the stimuli for the Nonlinguistic Visual condition, we translated Languages P and N into shapes (for a similar methodology, see Goldowsky, 1995). Each “word” was a single distinctive nonsense shape (e.g., a red asymmetric oval with yellow dots). Each shape was approx 3 in. in diameter. Category membership could not be induced by shape similarity, unlike in prior studies by Morgan and Newport (1981). The shapes were presented on a computer monitor, using SuperLab software running on a PowerPC. The shapes were presented, one at a time, in the center of the monitor, using the same timing parameters as those in the auditory experiments; presentation was sequential, with the onset of one shape following the offset of the previous shape. The Linguistic Visual condition was identical, except that instead of shapes, the nonsense words from Experiment 1 were presented in typed capital letters, one at a time, in the center of the monitor. Following exposure, participants received a forced-choice test analogous to the tests used in Experiments 1–3, in which they saw two sequences (of shapes in the Nonlinguistic Visual condition or of words in the Linguistic Visual condition). As in the auditory tasks, participants were asked to determine whether the first or the second sentence in the pair was more similar to the exposure language. Participants indicated their response via a key press.

Procedure

The procedure was identical to that of Experiment 3.

Results and Discussion

The first analysis asked whether subjects succeeded in learning Language P and Language N. For the Linguistic Visual condition, each group’s overall performance was significantly better than would be expected by chance: for Language P, the total score was 16.8 of a possible 24, $t(29) = 8.45$, $p < .0001$; for Language N, the total score was 15.9, $t(25) = 5.08$, $p < .0001$. Both groups’ overall performance was also significantly better than would be expected by chance for the Nonlinguistic Visual condition: for Language P, the total score was 17.9 of a possible 24, $t(25) = 13.38$, $p < .0001$; for Language N, the total score was 17.16, $t(24) = 10.01$, $p < .0001$. Table 3 presents subjects’ mean scores on the individual rules tested.

To assess differences in learning as a function of structural differences between the two languages, we submitted the overall scores for the two language groups in each condition to an ANOVA. In the Linguistic Visual condition, Language P and Language N learners did not differ, $F(1,54) = .82$, n.s. Similarly, in the Nonlinguistic Visual condition, Language P and Language N learners did not differ, $F(1,50) = 1.39$, n.s. Unlike in the auditory materials from Experiments 1–3, Language P was *not* easier for subjects to acquire than Language N when the materials were presented visually.

We applied the ANCOVA model from Experiment 1 to the results from each condition (see table 4). In the Nonlinguistic Visual condition, the only significant effects were Grammaticality [$F(1,90) = 118.66$] and First Word Legality [$F(1,90) = 5.09$]. The lack of a significant interaction between Grammaticality and Language [$F(1,90) = .003$, n.s.] is consistent with the results reported above, in which Language P and N scores did not differ. However, in the Linguistic Visual condition, both Grammaticality [$F(1,90) = 107.4$] and the Grammaticality \times Language interaction [$F(1,90) = 16.75$] were significant, as well as Similarity [$F(1,90) = 5.25$]. This result

indicates that while the Language P and N groups in the Linguistic Visual condition were not significantly different in the ANOVA reported above, removing the variance caused by other variables in the ANCOVA revealed an effect of predictive dependencies on performance. We hypothesize that the difference in results as a function of analytic technique may be due to increased sensitivity of the analysis of covariance. Participants may have used different strategies in this task. In particular, some participants may have verbalized the materials, essentially generating auditory materials despite the visual presentation. The increased sensitivity of the analysis of covariance may have permitted the discovery of subtle P versus N differences not apparent in the analysis of variance.

As in Experiment 3, we asked whether the linguistic and nonlinguistic visual materials were learned differently by contrasting the results of the Linguistic Visual and Nonlinguistic Visual conditions. The only difference between the two conditions lies in the materials, which were either nonsense shapes or written nonsense words. The grammars (Languages P and N) and test structures were identical. We submitted the total scores from the two conditions to a two-factor ANOVA including domain (linguistic versus nonlinguistic) and language (P versus N). None of the factors were significant: Language [$F(1,103) = 2.72$, n.s.]; Domain [$F(1,103) = 2.28$, n.s.]; interaction between Domain and Language [$F(1,103) = .18$, n.s.]. Thus, Language P and Language N were learned equivalently regardless of the linguistic status of the materials.

While the results from Experiment 4 support the conclusion from Experiment 3 that the linguistic status of the input does not affect learning in this task, these data suggest a possible modality effect. Unlike stimuli presented in the auditory domain, for which dependencies assist learners, the availability of predictive dependencies does not appear to affect learning in the visual domain to the same extent. This was not the result we predicted; we expected that if predictive dependencies afford greater learnability, this effect should be observed across presentation modalities. Instead, the results of Experi-

ment 4 stand in contrast with those of the auditory studies reported in Experiments 1–3.

Importantly, the findings from Experiment 4 support the contention that Language P is not *inherently* easier to learn than Language N or that the test invariably favors Language P learners. Instead, these results suggest that predictive dependencies impact learning in the auditory modality, but not the visual modality, at least for the stimuli used in these experiments. The absolute levels of performance are comparable across modalities; it is not the case that auditory learners outperform visual learners overall. What differ are the patterns of performance; visual and auditory presentations appear to elicit different constraints on learning, with a greater effect of dependency cues on auditory learning.

One reason the auditory and visual presentation conditions may have led to different outcomes concerns our original research question: do predictive dependencies assist learners in both linguistic and nonlinguistic tasks? It is possible that although the auditory nonlinguistic stimuli from Experiment 3 did not contain linguistic content—the “words” were beeps and buzzes taken from a bank of computer alert sounds—learners may have recoded the nonlinguistic sounds as linguistic. For example, listeners may have translated the sounds into words, encoding them as “high beep, chime, honk, burble . . .” If this is the case, then learners may have treated both of the auditory tasks as linguistic. Conversely, the visual tasks in Experiment 4 may have been treated as nonlinguistic. The nonsense shapes, which did not conform to known shapes or objects, were difficult to label linguistically. The mixed results for the nonsense words may reflect different processing strategies: some subjects may have processed the typed words linguistically, whereas others may have processed these materials as meaningless letter strings without linguistic content. We thus designed an additional experiment to attempt to replicate the modality differences observed in Experiments 1–4 using new stimuli. The materials in Experiment 5 were chosen so that the auditory stimuli would be difficult to label verbally, while the visual stimuli were easy to label. If ease of verbalization influenced

the apparent modality difference in Experiments 1–4, then the pattern of results should flip, such that the visual task should now show the effects of predictive dependencies. If, however, the original modality effect persists with the new stimuli, these findings would suggest that predictive dependencies affect learning of sequential stimuli in auditory tasks, but not visual tasks.

EXPERIMENT 5

This study is a replication of the auditory nonlinguistic task used in Experiment 3 and the visual nonlinguistic task used in Experiment 4. We used a new set of nonlinguistic sounds that are difficult to label—various types of drums and bells—and a new set of nonlinguistic shapes that are easy to label—familiar shapes, such as circles, triangles, and hearts. Following exposure, learners received the test used in the previous experiments, implemented in the vocabulary of sounds or shapes used during exposure. If the modality effect observed in the prior experiments was an artifact of stimulus choice or ease of labeling, then learners acquiring visual stimuli should now be more affected by the presence or the absence of predictive dependencies. If, however, the original modality effect persists, then learners in the auditory condition, but not the visual condition, should show enhanced performance on Language P relative to Language N.

Method

Participants

One hundred and twelve monolingual English speaking undergraduates at the University of Wisconsin-Madison participated in this study for course extra credit. Fifty-two of the participants were assigned to the Nonlinguistic Auditory condition, and 60 were assigned to the Nonlinguistic Visual condition. Within each condition, participants were assigned to either Language P or Language N.

Materials

The Nonlinguistic Auditory stimuli consisted of digitized recordings of various types of bells and drums. Each sound corresponded to a word

from the vocabulary. The sounds were presented using the procedures from Experiment 3. The stimuli for the Nonlinguistic Visual condition consisted of familiar shapes such as circles, squares, triangles, and crosses. The shapes were presented using the procedures from Experiment 4. Following exposure, participants received a forced-choice task analogous to the tests used in the previous experiments: learners either saw two sequences of shapes (in the Nonlinguistic Visual condition) or heard two sequences of sounds (in the Nonlinguistic Auditory condition) and judged which sequence was more similar to the stimuli observed during exposure.

Procedure

The procedure was identical to those for Experiment 3 (for auditory stimuli) and Experiment 4 (for visual stimuli).

Results and Discussion

The first analysis asked whether subjects succeeded in learning Language P and Language N. For the Nonlinguistic Auditory condition, each group's overall performance was significantly better than would be expected by chance: for Language P, the total score was 17.0 of a possible 24, $t(26) = 9.37$, $p < .0001$; for Language N, the total score was 14.4, $t(24) = 3.34$, $p < .01$. Each group's overall performance was also significantly better than would be expected by chance for the Nonlinguistic Visual condition: for Language P, the total score was 17.04 of a possible 24, $t(23) = 5.77$, $p < .0001$; for Language N, the total score was 16.77, $t(35) = 12.05$, $p < .0001$. Mean scores on the individual rules are shown in Table 3.

To assess differences in learning as a function of structural differences between the two languages, we submitted the overall scores for the two language groups in each condition to an ANOVA. In the Nonlinguistic Auditory condition, Language P learners significantly outperformed Language N learners, $F(1,50) = 8.74$, $p < .01$. As in the auditory presentations from Experiments 1–3, participants listening to the input performed better given Language P than Language N. However, in the Nonlinguistic Visual condition, Language P and N learners

did not differ, $F(1,58) = .09$, n.s. As in Experiment 4, Language P was *not* easier for subjects to acquire than Language N when presented visually. To ensure that this pattern of results was not due to surface variables, we applied the ANCOVA model from Experiment 1 to these data, as shown in Table 4. For the Nonlinguistic Auditory condition, the only significant effects were Grammaticality [$F(1,90) = 26.4$] and the Grammaticality \times Language interaction [$F(1,90) = 9.35$], supporting the hypothesis that the differential performance of the Language P and N groups was due to structural properties of the two languages. For the Nonlinguistic Visual condition, only the main effect of Grammaticality was significant [$F(1,90) = 44.75$]; predictive dependencies did not affect performance in this condition.

Visual versus Auditory Results

We next compared the two conditions from Experiment 5 to one another to determine

whether modality of presentation affected the results. The main effect of Language (P versus N) [$F(1,108) = 6.67$, $p < .05$] and the interaction between Language and Modality [$F(1,108) = 4.63$, $p < .05$] were both significant, while the main effect of modality (visual versus auditory) [$F(1,108) = 3.01$, n.s.] was not significant. The significant interaction suggests that the effects of predictiveness were not uniform across modalities and that, consistent with the separate condition analyses, Language P was easier to acquire than Language N only when the presentation was auditory. These findings replicate the pattern of results observed across the first four experiments.

Overall Analyses

To further explore the locus of effects across experiments, the next set of analyses compared the results from Experiments 1–5, as shown in Fig. 1. The $2 \times 2 \times 2$ ANOVA contrasted language (P versus N), linguistic status (linguistic

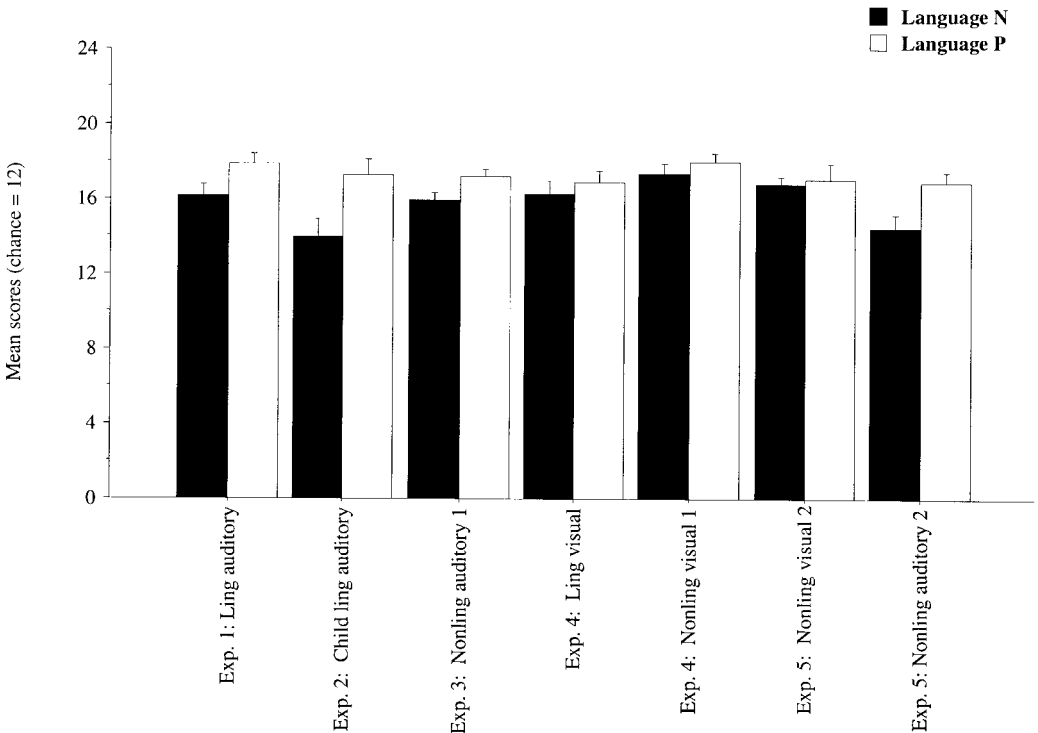


FIG. 1 Mean scores and standard errors for the Language P and N groups for Experiments 1–5.

vs. nonlinguistic), and modality (visual versus auditory). The main effect of Language was significant, with Language P learners outperforming Language N learners: $F(1,354) = 17.28$, $p < .0001$. None of the other main effects were significant: Linguistic status $F(1,354) = .79$, n.s.; Modality $F(1,354) = 3.38$, n.s. Only the interaction between Language and Modality was significant, $F(1,354) = 4.98$, $p < .05$; all other interactions, $F(1,266) < 2$, n.s. Consistent with the results of the individual experiments, presentation modality affected the degree to which the availability of predictive cues affected the results, with Language P performance exceeding Language N performance only in the auditory conditions. Importantly, whether the materials were linguistic or nonlinguistic did not affect the results, supporting the hypothesis that a constraint to detect predictive dependencies is not tied solely to language learning.

Modality Effects

Experiments 1–5 revealed interesting differences between sequential learning in the auditory and visual domains. While all of the experiments using auditory materials elicited stronger performance on Language P than on Language N, the experiments using visual materials revealed no differences between the languages. The overall levels of performance were comparable across modalities, consistent with prior findings that basic statistical learning processes, such as detecting transitional probabilities, operate similarly across domains (e.g., Fiser & Aslin, 2001; Hunt & Aslin, 2001; Saffran et al., 1996b, 1999; Saffran & Griepentrog, 2001). The modality differences appear to arise when we consider the impact of predictive dependencies which, unlike the transitional probabilities explored in our previous work, are computed over word categories (rather than individual tokens) and which generate hierarchical relationships not tied to immediate adjacencies.

Why might predictive dependencies influence sequence learning in the auditory domain, but not in the visual domain? One hypothesis is that predictive relationships among items presented sequentially are processed preferentially in audition due to the generally sequential na-

ture of the auditory world. Auditory information is fleeting by nature; sounds do not persist in time. This is most obviously true of linguistic information, where sounds occur in rapid succession, requiring the listener to integrate over a window of time. Most other auditory experiences are similarly sequenced: for example, consider musical patterns, nonlinguistic vocalizations across species, and passing footsteps. The nature of the auditory world requires listeners to track sequences and to note the relationships between events separated in time. Visual processing also contains a temporal aspect, but the visual world is typically more stable and less fleeting than the auditory world. Interrogating a visual scene requires the viewer to track the relationships of objects in space and to note spatiotemporal correlations between parts of objects to detect movement, but unlike with audition, the objects themselves persist in time. The processing capacity called upon by visual scenes thus entails simultaneous processing of information in the viewer's environment, leading to the speculation that visual information is inherently less sequential than auditory information (with notable exceptions, such as signed languages, gesture, and facial expressions). If this is the case, then materials in the visual modality may not tap into a constraint to use the predictiveness of elements to acquire sequential structure to the same extent as the processing of auditory information.

These differences between the auditory and visual environments are consistent with the oft-cited observation that learners in serial recall tasks actually perform better given auditory than linguistic stimuli (see Penney, 1989, for extensive review). Modality effects indicating auditory superiority for tasks requiring sequential learning and memory appear across an array of procedures, including short-term memory tasks with linguistic and nonlinguistic materials, order judgments, frequency estimation, rhythm perception, suffix effects, temporal output order, and even the resolution of temporal anaphors (e.g., Broadbent, 1956; Frick, 1985; Glenberg & Fernandez, 1988; Glenberg & Jona, 1991; Jakimik & Glenberg, 1990; Penney, 1975; Rollins, Schurman, Evans, & Knoph,

1975; Savin, 1967; Watkins & Peynircioglu, 1983).

Interestingly, visual superiority emerges when tasks entail simultaneous processing rather than sequential processing (e.g., Broadbent, 1956; Penney, 1989; Rollins et al., 1975). This literature is consistent with the observation that what must be learned in the visual environment often requires attention to simultaneously present elements arrayed in space. It is possible that, given a visual task that entailed simultaneously present predictive dependencies rather than sequential dependencies, learners would show the same type of Language P advantage as we found in the auditory experiments using sequential stimuli. We designed Experiment 6 to test the hypothesis that learners engaged in visual tasks use predictive dependencies between elements simultaneously present in the display. That is, unlike the sequential presentation used in the previous experiments, learners in a visual task might capitalize on the predictive dependencies in Language P given simultaneous presentations.

EXPERIMENT 6

This experiment was a conceptual replication of the Nonlinguistic Visual condition from Experiment 4. Rather than presenting the shapes one by one, with the same timing parameters as in the auditory experiments, each "sentence" in Experiment 6 was presented simultaneously, with all of the shapes from the sentence arrayed spatially on the screen for 3 s. Predictiveness in the simultaneous task entailed the same pattern of dependencies as in the sequential task, but without respect to sequential order. For example, in Language P, if a D word occurred on the screen, an A word simultaneously occurred on the screen. However, in Language N, a D word could occur either with or without an A word. Other than the simultaneity of presentation, Experiment 6 was identical to Experiment 4 in the shapes and sentences used during exposure and testing. We hypothesized that learners might be attuned to dependencies between visual elements when those elements are simultaneously available, leading to a Language P advantage.

Method

Participants

Fifty-six monolingual English speaking undergraduates at the University of Wisconsin-Madison participated in this study for course extra credit. Half of the participants were assigned to Language P and half were assigned to Language N.

Materials

The vocabulary was drawn from the Nonlinguistic Visual condition from Experiment 4, which consisted of distinctive nonsense shapes. The shapes were presented on a computer monitor, using SuperLab software running on a PowerPC. Each sentence (consisting of three to five shapes) was displayed on the monitor, with shapes arrayed such that each form class always occurred in a particular position on the screen. That is, "A word" shapes always occurred in the upper righthand corner, whereas "F word" shapes always occurred in the center of the bottom of the screen. Each shape sentence was shown for 3 s, with a 2-s blank screen between sentences. We chose to use this arrayed layout, rather than a sequentially ordered layout, to decrease the probability that learners would use a left-to-right sequential processing strategy. Following exposure to either Language P or N, participants received a forced-choice test analogous to the tests used in Experiments 1–5, in which they saw two shape sentences, each arrayed spatially. Participants were asked to determine whether the first or the second sentence in the pair was more similar to the exposure language. Participants indicated their response via a key press.

One difference in the test from the previous experiments concerns Rule 2. Because Rule 2 tests knowledge of a shift in sequential position (flipping the positions of D and G), items testing this rule necessarily differed from those used in the sequential tasks. Instead of switching the temporal positions of D and G words, we switched the spatial positions of D and G words. For example, if during exposure, D words occurred in the top right corner and G words occurred in the center of the screen,

these positions were switched for ungrammatical items testing Rule 2. Because of this change, Rule 2 no longer assessed anything about the grammatical structure of the language; instead, Rule 2 assessed whether learners remembered the spatial position of individual elements. We thus collected Rule 2 data so that the test was equal in duration to the tests used in Experiments 1–5, but did not include the Rule 2 data in the analyses, as these data are not pertinent to the acquisition of grammar. We did not substitute additional rules because this would have complicated comparisons with the prior experiments.

Procedure

Other than the simultaneous presentation of the shape sentences, the procedure was identical to those for Experiments 4 and 5 (visual condition).

Results and Discussion

The first analysis asked whether subjects succeeded in learning Language P and Language N. Both groups performed significantly better than would be expected by chance (total scores = Rules 1, 3, and 4): for Language P, the total score was 15.39 of a possible 18, $t(27) = 17.52$, $p < .0001$; for Language N, the total score was 13.29, $t(27) = 8.69$, $p < .0001$. Table 3 presents subjects' mean scores on the individual rules tested.

To assess differences in learning as a function of structural differences between the two languages, we submitted the overall scores for the two language groups (for Rules 1, 3, and 4)³ to an ANOVA. Language P learners significantly outperformed Language N learners, $F(1,54) = 11.76$, $p < .01$. These findings support the hypothesis that learners detect and use predictive dependencies in visual tasks when the stimuli are presented simultaneously. To ensure that these results were not due to surface variables,

the data were submitted to an ANCOVA. Several of the covariates used in the prior analyses were not applicable, given the lack of sequential information in these materials; the model thus included only the Grammaticality and Language factors and the Length and Similarity covariates. As shown in Table 5, the main effects of Grammaticality [$F(1,90) = 251.82$] and Similarity [$F(1,90) = 4.03$] were significant, as was the Grammaticality \times Language interaction [$F(1,90) = 4.63$]. These results suggest that the differential performance of the Language P and N groups was not due to surface variables, but was a function of the availability of predictive dependencies in the input.

We next compared the results from Experiment 6 to the results from the analogous condition of Experiment 4, the Nonlinguistic Visual condition, in which the same shapes were used, but sentences were presented sequentially. The ANOVA included two factors: Language (P versus N) and Mode of presentation (sequential versus simultaneous). The dependent variable was the mean score including Rules 1, 3, and 4 (as discussed above, the use of simultaneous presentation in Experiment 6 altered what Rule 2 was testing, making it difficult to compare performance on this rule across experiments). Both of the main effects were significant: Language [$F(1,101) = 11.18$, $p < .01$]; Mode [$F(1,101) = 4.63$, $p < .05$]. This pattern of results suggests that Language P learners outperformed Language N learners overall and that learners exposed to material in the simultaneous mode outperformed learners exposed to material in the sequential mode. The interac-

TABLE 5
ANCOVA *F*-Values for Experiment 6

Factor	Experiment 6 Simultaneous Visual
Grammaticality	251.8**
Language	0.08
Grammaticality \times Language	4.63*
Length	0.09
Similarity	4.03*

* $p < .05$.

** $p < .01$.

df = 1, 90.

³Analyses including Rule 2 show the same pattern of results as the reported analyses excluding the Rule 2 data: $F(1,53) = 8.31$, $p < .01$. Because Rule 2 did not assess acquisition of the grammar given simultaneous presentation, we focus here on the results excluding Rule 2.

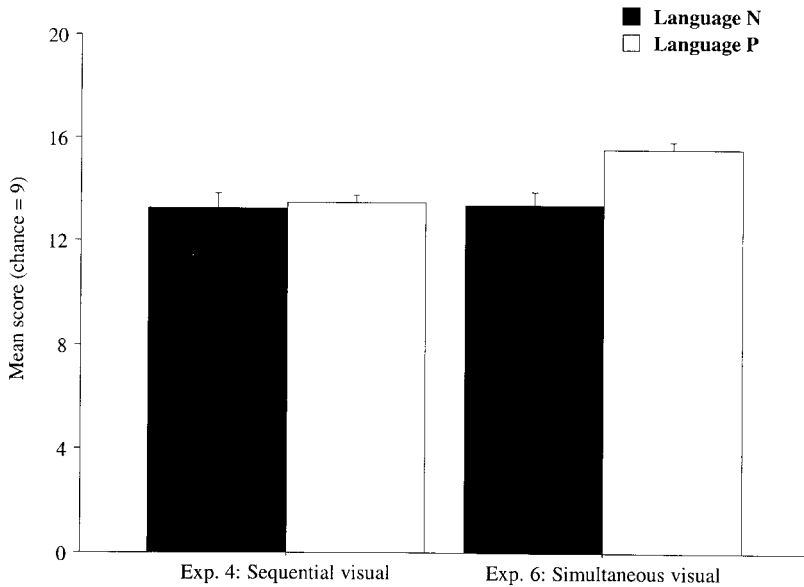


FIG. 2 Mean scores and standard errors for the language P and N groups for the sequentially presented shapes in Experiment 4 (Nonlinguistic Visual condition) and the simultaneously presented shapes in Experiment 6.

tion between Language and Mode was also significant, [$F(1,101) = 4.89, p < .05$]. This result suggests that the mode of presentation had a differential effect on the use of predictive dependencies. As shown in Fig. 2, learners in the sequential condition (Experiment 4, Nonlinguistic Visual condition) showed no difference in learning rates as a function of the availability of predictive dependencies. However, predictive dependencies did affect learning in the simultaneous condition (Experiment 6). Relative to the other three groups, the learners in the simultaneous Language P group performed best. It is unclear whether this was due to positive effects of predictive dependencies on learning in Language P or deleterious effects of the absence of predictive dependencies in Language N; the baseline level of performance in simultaneous visual tasks may exceed the baseline level of performance in sequential visual tasks. Nevertheless, the results support the hypothesis that learning in the visual system is more attuned to dependencies between simultaneously available elements arrayed in space than to sequentially available elements arrayed in time.

GENERAL DISCUSSION

This series of experiments was designed to address two questions. First, do human learners detect and use predictive dependencies, like those that characterize phrases in natural languages, as a cue to linguistic structure? Second, is the use of predictive dependencies reserved solely for linguistic tasks, or does this learning mechanism operate in nonlinguistic domains as well? The results of Experiment 1 suggest that adults were more successful at learning an artificial language when the grammar includes predictive dependencies as a cue to phrase structure. Experiment 2 extended these results to include child learners, suggesting a constraint on learning that may be available during the years in which children acquire their native language. Experiment 3 demonstrated that the use of predictive dependencies in learning phrase structure is not limited to language learning tasks.

While the effect of predictive dependencies reliably emerged across these experiments, the differences in performance across language groups were not large. In particular, Language N learners were quite successful overall, though

not quite as successful as the Language P learners. It is possible that a more sensitive test might show greater differences. The test used in these experiments does not invariably target knowledge of the underlying structure of the language—learners could succeed on many test items by knowing something about which items go where and which items follow which other items. It is thus possible that a test assessing deeper structure knowledge (perhaps involving transformations) would tease the two language groups' performance apart to a greater extent. In addition, the generality of our conclusions is limited by the use of only a single pair of grammars; it would be extremely useful to examine predictive dependencies in other types of structures, including grammars like English (and many nonlinguistic systems, such as music) in which dependencies link events in a forward direction, unlike the backward dependencies used here. Moreover, the dependencies tested in these grammars were limited to neighboring elements, unlike the long-distance dependencies characterizing natural languages. Nevertheless, the results point to a possible constraint on learning: humans can detect and use predictive dependencies to acquire phrase structure and perform more poorly when these dependencies are not present amongst the statistics of the input, without respect to the linguistic nature of the task.

Modality Effects

The results of Experiments 4–6 support the hypothesis that predictive dependencies are used by learners when the dependencies lie between elements presented in a manner appropriate to perceptual learning capacities in each modality. In the auditory modality, where information is generally serial and fleeting, sequential presentation elicits effects of predictive dependencies: dependencies allow learners to tie together events across time. Learners can also detect and use predictive dependencies in the visual modality, but not when the input is sequential. Instead, learners make use of dependencies when they link spatially arrayed and simultaneously available elements.⁴ It is unclear whether these effects are due to inherent per-

ceptual/processing differences or to experience in each modality. That is, the perceptual learning systems in each modality may be specialized to perform best on sequential or simultaneous input, in the absence of experience. Alternatively, our processing capacities in each modality may have been shaped via experience to specialize in different types of learning, given the nature of the auditory and visual worlds. One way to explore these two explanations for the modality difference would be to test individuals who have been extensively exposed to a signed language. Signed languages contain extensive sequential structure (in addition to simultaneous structure), with elements that must be tracked and combined over time as in spoken languages. It is possible that individuals who sign would be sensitive to the predictive dependencies in the visual experiments, since they may be more specialized in detecting and using sequential structure in the visual modality than individuals who are not speakers of signed languages. Such manipulations would allow us to tease apart the cause of the visual/auditory modality differences in the use of predictive dependencies given sequentially presented input.

Constraints on Statistical Learning

The predictive dependencies internal to phrases are a hallmark of natural languages. However, organization into phrases and hierarchies also characterizes nonlinguistic sequential information processing (e.g., Lashley, 1951). The kinds of structure at issue serve to organize and package serial information into manageable chunks, which then enter relationships with one another. The generation of hierarchical structure presumably maximizes cognitive economy, facilitating the transmission of more complex information than could otherwise be transmitted in a serial channel. 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

⁴It is unclear how to devise a simultaneous auditory language to fill out the parametric variations in modality \times simultaneous/sequential presentation.

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 the packaging of serial inputs into higher order organization facilitates not only language production and processing, but also language acquisition. Systems that are highly organized are more learnable than systems that are not—as long as the system of organization is consistent with the learner’s cognitive structure. These ideas suggest a possible 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 (e.g., Bever, 1970; Christiansen, 1994; Christiansen & Devlin, 1997; Ellefson & Christiansen, 2000; Newport, 1982, 1990). Applying these ideas to the current research, the predictive dependencies that characterize phrase structure may recur cross-linguistically because they enhance learnability. On this view, languages evolve to fit the human learner. To the extent that this type of view is correct, the striking similarities among human languages may reflect 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 unit, the phrase, we find the beginnings of an explanation for why languages contain within-phrase dependencies: human learners may best acquire internal structure in sequential input when that structure is marked by strong predictive relationships between elements. 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, computational research

suggests that universal word order typologies may 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—or any other type of theory that assigns an important role to linguistic input—are necessarily under-constrained. As research on animal learning has amply demonstrated, learning in biological systems is highly constrained (e.g., Gallistel, 1990; 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. More generally, our focus on learning provides a needed bridge between theories focused on nature and theories focused on nurture, because constrained learning mechanisms require both experience to drive learning and preexisting structures to capture and manipulate those experiences.

APPENDIX 1:

Language P Sentences

1. ACF

biff cav dupp
 hep lum loke
 mib neb jux
 rud sig vot
 biff lum dupp
 hep cav jux
 mib sig loke
 rud neb vot

2. ADCF

biff klor lum dupp
 hep pell neb loke
 mib klor sig jux
 rud pell cav vot
 hep klor sig dupp
 biff pell sig vot
 mib pell lum jux
 rud klor cav loke

3. ACGF

biff cav tiz dupp
 hep lum pilk loke
 mib neb tiz jux

APPENDIX 1—*continued*

rud sig pilk vot
 rud neb pilk dupp
 mib lum tiz loke
 biff cav pilk jux
 hep sig tiz vot

4. ADCGF
 biff klor cav pilk jux
 biff pell sig tiz vot
 rud pell lum tiz dupp
 mib klor lum pilk loke
 mib pell cav tiz jux
 rud klor sig pilk vot
 hep klor neb tiz loke
 rud pell neb pilk dupp

5. ACFC
 biff sig dupp cav
 hep cav loke neb
 mib lum jux sig
 rud neb vot lum
 rud cav jux lum
 hep sig loke neb

6. ADCFC
 mib klor cav vot sig
 rud pell lum loke neb
 hep klor sig dupp lum
 biff pell neb jux cav

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

8. ACFCG
 biff neb jux lum tiz
 hep cav loke neb pilk
 mib sig dupp cav pilk
 rud lum vot sig tiz

Language N Sentences

1. ACF
 biff cav dupp
 hep lum loke
 mib neb jux
 rud sig vox

2. ADCF
 bif klor lum dupp
 hep pell neb loke
 mib klor sig jux
 rud pell cav vot
 hep klor sig dupp
 biff pell neb vot

3. DCF
 klor neb jux

klor sig dupp
 pell cav vot

4. AGF
 biff tiz jux
 hep pilk lok
 mib tiz loke
 rud pilk vot

5. ADGF
 biff pell tiz dupp
 mib pell pilk jux
 rud klor tiz loke
 hep klor pilk vot
 rud pell tiz jux
 mib klor pilk dupp
 hep pell tiz vot

6. DGF
 klor pilk loke
 klor pilk dupp

7. ACGF
 biff cav tiz dupp
 hep lum pilk loke
 mib neb tiz jux
 rud sig pilk vot
 rud neb pilk dupp
 mib lum tiz loke
 biff neb pilk jux
 hep cav pilk vot

8. ADCGF
 biff klor cav pilk jux
 biff pell sig tiz vot
 rud pell lum tiz dupp
 mib klor lum pilk loke
 mib pell cav tiz jux
 rud klor sig pilk vot
 hep klor neb tiz loke

9. DCGF
 klor neb pilk jux
 pell lum pilk dupp
 klor sig tiz vot
 pell cav tiz loke
 klor neb tiz jux
 pell sig pilk dupp
 klor lum tiz vot
 pell sig tiz loke
 rud pell cav pilk dupp

APPENDIX 2:

Test Items

Rule 1: Sentences Must Contain an A Phrase

biff klor sig pilk jux	[A-D-C-G-F]
*sig pilk jux	[C-G-F]
hep pell lum tiz dupp	[A-D-C-G-F]
*lum tiz dupp	[C-G-F]

APPENDIX 2—*continued*

mib klor cav tiz vot	[A-D-C-G-F]
*cav tiz vot	[C-G-F]
rud pell neb pilk loke	[A-D-C-G-F]
*neb pilk loke	[C-G-F]
biff pell sig pilk dupp	[A-D-C-G-F]
*sig pilk dupp	[C-G-F]
hep klor neb tiz dupp	[A-D-C-G-F]
*neb tiz dupp	[C-G-F]

Rule 2: D Words Follow A Words, while G Words Follow C Words

biff klor lum pilk jux	[A-D-C-G-F]
*biff pilk lum klor jux	[A-G-C-D-F]
hep pell cav pilk dupp	[A-D-C-G-F]
*hep pilk cav pell dupp	[A-G-C-D-F]
mib klor sig tiz vot	[A-D-C-G-F]
*mib tiz sig klor vot	[A-G-C-D-F]
rud pell neb pilk loke	[A-D-C-G-F]
*rud pilk neb pell loke	[A-G-C-D-F]
mib pell cav tiz dupp	[A-D-C-G-F]
*mib tiz cav pell dupp	[A-G-C-D-F]
rud klor lum pilk vot	[A-D-C-G-F]
*rud pilk lum klor vot	[A-G-C-D-F]

Rule 3: Sentences Must Contain an F Word

biff klor neb loke	[A-D-C-F]
*biff klor neb	[A-D-C]
mib lum pilk jux	[A-C-G-F]
*mib lum pilk	[A-C-G]
hep klor cav tiz vot	[A-D-C-G-F]
*hep klor cav tiz	[A-D-C-G]
rud pell sig tiz dupp	[A-D-C-G-F]
*rud pell siz tiz	[A-D-C-G]
biff pell sig jux	[A-D-C-F]
*biff pell sig	[A-D-C]
hep neb tiz vot	[A-C-G-F]
*hep neb tiz	[A-C-G]

Rule 4: C Phrases Must Precede F Words

rud pell neb dupp	[A-D-C-F]
*rud pell dupp	[A-D-F]
mib klor cav jux	[A-D-C-F]
*mib klor jux	[A-D-F]
hep klor lum vot	[A-D-C-F]
*hep klor vot	[A-D-F]
hep pell sig pilk loke	[A-D-C-G-F]
*hep pell loke	[A-D-F]
hep pell neb pilk jux	[A-D-C-G-F]
*hep pell jux	[A-D-F]

mib klor sig tiz loke	[A-D-C-G-F]
*mib klor loke	[A-D-F]

REFERENCES

- 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: Holt.
- Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon Press.
- 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., & 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.
- Ellefsen, M. R., & Christiansen, M. H. (2000). Subjacency constraints without universal grammar: Evidence from artificial language learning and connectionist modeling. In *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
- 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.
- Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, **12**, 499–504.
- Frick, R. W. (1985). Testing visual short-term memory: Simultaneous versus sequential presentations. *Memory & Cognition*, **13**, 346–356.
- Gallistel, C. R. (1990). *The organization of learning*. MIT Press: Cambridge, MA.
- 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 & L. R. Gleitman (Eds.), *Language acquisition: The state of the art*. Cambridge: Cambridge Univ. Press.
- Glenberg, A. M., & Fernandez, A. (1988). Evidence for auditory temporal distinctiveness: Modality effects in order and frequency judgments. *Journal of Experimental Psychology: Learning Memory and Cognition*, **14**, 728–739.
- Glenberg, A. M., & Jona, M. (1991). Temporal coding in rhythm tasks revealed by modality effects. *Memory & Cognition*, **19**, 514–522.
- Goldowsky, B. (1995). *Learning structured systems from imperfect information*. Unpublished Ph.D. dissertation, University of Rochester.

- Gómez, R. L., & Gerken, L. A. (1999). Artificial grammar learning by one-year-olds leads to specific and abstract knowledge. *Cognition*, **70**, 109–135.
- 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: Univ. of Chicago Press.
- Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Simultaneous extraction of multiple statistics. *Journal of Experimental Psychology: General*, **130**, 658–680.
- Jakimik, J., & Glenberg, A. (1990). Verbal learning meets psycholinguistics: Modality effects in the comprehension of anaphora. *Journal of Memory and Learning*, **29**, 582–590.
- 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.
- 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.
- MacWhinney, B. (1999). *The emergence of language*. Mahwah, NJ: Erlbaum.
- Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, **283**, 77–80.
- 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.
- McAndrews, M. P., & Moscovitch, M. (1985). Rule-based and exemplar-based classification in artificial grammar learning. *Memory & Cognition*, **19**, 469–475.
- 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.
- Muelemans, T., & Van der Linden, M. (1997). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, **23**, 1007–1028.
- Newport, E. L. (1982). Task specificity in language learning? Evidence from speech perception and American Sign Language. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art*. Cambridge: Cambridge Univ. Press.
- Newport, E. L. (1990). Maturation constraints on language learning. *Cognitive Science*, **14**, 11–28.
- Penney, C. G. (1975). Modality effects in short-term verbal memory. *Psychological Bulletin*, **82**, 68–84.
- Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, **17**, 398–422.
- 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.
- 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.
- 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.
- Rescorla, R. A. (1966). Predictability and number of pairings in Pavlovian fear conditioning. *Psychonomic Science*, **4**, 383–384.
- Rollins, H. A., Schurman, D. L., Evans, M. J., & Knoph, K. (1975). Auditory versus visual processing of three sets of simultaneous digit pairs. *Journal of Experimental Psychology: Human Learning and Memory*, **104**, 173–181.
- Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, **44**, 493–515.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996a). Statistical learning by 8-month-old infants. *Science*, **274**, 1926–1928.
- Saffran, J. R., & Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: Evidence for developmental reorganization. *Developmental Psychology*, **37**, 74–85.
- 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. (1996b). 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.
- Savin, H. B. (1967). On the successive perception of simultaneous stimuli. *Perception & Psychophysics*, **2**, 479–482.
- Seidenberg, M. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science*, **275**, 1599–1603.
- Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, **23**, 569–588.
- Servan-Schreiber, E., & Anderson, J.R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **16**, 592–608.
- 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.
- 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.
- Watkins, M. J., & Peynircioglu, Z. F. (1983). Interaction between presentation modality and recall order in memory span. *American Journal of Psychology*, **96**, 315–322.

(Received April 11, 2001)

(Revision received July 20, 2001)