LANGUAGE LEARNING AND DEVELOPMENT, *3*(1), 1–42

Statistical Learning of Syntax: The Role of Transitional Probability

Susan P. Thompson
Hobart and William Smith Colleges

Elissa L. Newport
University of Rochester

Previous research has shown that, for learners to fully acquire a miniature phrase structure language, the language must contain cues to the phrases—for example, prosodic grouping or morphological agreement of the words within a phrase (Morgan, Meier, & Newport, 1987, 1989). Research on word segmentation has shown that learners can use transitional probabilities between syllables to segment speech into word-like units (Saffran, Aslin, & Newport, 1996). In the present research, we combine and extend these two sets of findings, asking whether learners can use transitional probabilities between *words* (or word classes) to segment sentences into *phrases,* and use this phrasal information to fully acquire the syntax of a miniature language.

Adult subjects were exposed to sentences from a miniature language. A pattern in the transitional probabilities between words—high within phrases, low at phrase boundaries—was created by adding syntactic properties that are widespread in natural languages: optional phrases, repeated phrases, moved phrases, different-sized form classes, or all four properties combined. All conditions outperformed controls in learning the language. The best learning occurred with all properties combined, despite the fact that this language was the most complex. These data address the important question of how language learning is successful in the face of the massive complexity of natural languages. In our experiments, learning got better, not worse, when *properly structured* complexity was added to a language. The results also show that the same type of statistical computation useful in word segmentation might be used as well in learning syntax, suggesting that the range of statistics needed for acquiring various types of structure in natural languages might be suitably small.

Correspondence should be addressed to Susan P. Thompson, Department of Psychology, 205 Gulick Hall, Hobart and William Smith Colleges, Geneva, NY 14456. E-mail: thompson@hws.edu

An important question in the field concerns the mechanisms by which human learners acquire the syntax of their native language. Researchers from all perspectives suggest that this process requires contributions from both nurture and nature—that is, from both the linguistic environment to which learners are exposed and some innate predispositions of human learners to process and learn temporally organized patterns in particular ways (for discussion, see Chomsky, 1965; Gleitman & Newport, 1995; Lidz, Gleitman, & Gleitman, 2003; Marcus, 2001; Pinker, 1994; Seidenberg, 1997). However, little is known about the precise processes by which this learning occurs or the mechanisms responsible for its rapidity and success.

In recent work a number of investigators have begun to examine the role that distributional or statistical approaches to learning might play in the acquisition of language, including the acquisition of phonology (Maye & Gerken, 2001; Maye, Werker, & Gerken, 2002), word segmentation (Newport & Aslin, 2000, 2004; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996), form class (Hunt, 2002; Mintz, Newport, & Bever, 1995, 2002), and syntax (Gerken, Wilson, & Lewis, 2005; Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005; Gomez, 1997, 2002; Gomez & Gerken, 1999; Saffran, 2001, 2002). In the present article, we ask whether a particular statistical learning approach that we have proposed for the problem of word segmentation—involving the computation of transitional probabilities between sequential elements—might be extended to facilitate the acquisition of syntax.

PREVIOUS APPROACHES TO THE PROBLEM OF SYNTAX ACQUISITION

Investigators agree that the task of acquiring the syntax of a natural language is a difficult and complex problem, one for which constraints and biases that focus and direct early learning will be necessary. Most previous approaches have suggested some type of “bootstrapping” mechanism that might be utilized to direct early learning. One such approach (though not conventionally called “bootstrapping”) hypothesizes that children come to the task of syntax acquisition with extensive innate knowledge about the principles by which natural language grammars are universally organized and the parameters along which the grammars of specific languages will vary. This innate knowledge then greatly constrains the analyses that children perform to learn their native language (Chomsky, 1957, 1965, 1981, 1995). A different type of approach involves using other, more accessible cues than syntax itself to acquire the complex syntax indirectly. For example, “semantic bootstrapping” (Bowerman, 1973; Pinker, 1984) hypothesizes that children may use meaning to acquire the word classes and their ordering; “prosodic bootstrapping” (Gleitman & Wanner, 1982; Morgan, 1986; Morgan & Newport, 1981; Morgan et al., 1987) hypothesizes that children may use the prosodic grouping of words to learn their analysis into phrases.

However, recent research on statistical approaches to language acquisition has provided surprising evidence about young children's ability to perform distributional analyses of many aspects of linguistic input during rapid online processing. This research has shown that adults, infants, and young children can acquire the structure of words and word classes by analyzing the consistency of sequences that occur in the auditory input and by using the contrast between highly consistent and less consistent sequences to distinguish words and categories from accidental co-occurrences. Here we ask whether analyses of word or word class sequences might account for parts of syntax acquisition as well.

MINIATURE ARTIFICIAL GRAMMARS AND THE ACQUISITION OF SYNTAX THROUGH DISTRIBUTIONAL ANALYSES

The use of miniature artificial grammars to study language learning has a long history in the literature (e.g., Braine, 1963; Reber, 1967). A question of continuing interest concerns how learners acquire the permissible sequences of words and word classes within sentences. In one early study, Smith (1969) showed that adult learners of artificial languages displayed surprising limits in their ability to acquire word-class dependencies. He presented adult learners with two-word strings and found that they learned the serial position of words, but not dependencies between the word classes. In response to this finding, Moeser and Bregman (1972) suggested that semantic referents are necessary mediators in the acquisition of syntax. They exposed participants to a miniature phrase structure grammar with a matching reference world. Their participants learned complex aspects of syntax only to the extent that these were correlated with similar relationships in the physical reference world.

However, Morgan and Newport (1981) suggested that what was crucial for complex syntax acquisition was not semantics per se, but more generally a set of cues from which the phrase structure of the language might be acquired. The goal of the acquisition of syntax, of course, is to acquire the order of words within sentences in a particular language. However, in all natural languages, the order in which the words may occur is best captured through a hierarchical description: words are ordered in various ways to form phrases, and phrases are ordered in various ways to form sentences (Chomsky, 1957; Jackendoff, 1977). Morgan and Newport (1981) showed that the reference world of Moeser and Bregman (1972) was successful in facilitating the acquisition of complex aspects of syntax because it served to demarcate the phrasal groupings of the language. In further studies, Morgan et al. (1987, 1989) showed that there are many cues to phrase structure, including local cues (such as prosody and function words) and cross-sentential cues (such as pronominalization and permutation), that are similarly effective in leading the learner to induce both linguistic dependencies between classes of words and the hierarchical phrase structure of the language.

All of the cues tested in Morgan et al. (1987, 1989) were *in addition* to the basic phrase structure of the language. Saffran (2001) argued that there are cues *within* a basic phrase structure—namely, the predictive relationships between classes of words—that can serve as cues to phrase structure as well. She used the same language as Moeser and Bregman (1972), Morgan and Newport (1981), and Morgan et al. (1987, 1989), but altered it slightly so that there were perfect dependencies between classes of words within two of the three phrases of the language. Saffran's participants scored significantly above chance, indicating that learners are sensitive to these predictive relationships.

However, although these studies suggested general features of language that contribute to the learning of syntactic phrase structure—semantics, dependencies between classes of words, and the like—none suggested a precise mechanism through which the learning occurred. This was first addressed in a different domain of language acquisition, namely, word segmentation. Saffran, Aslin, and Newport (1996) showed that 8-month-old infants possess a powerful learning mechanism capable of segmenting a fluent speech stream into wordlike units after only 2 min of listening exposure. Saffran, Newport, and Aslin (1996) suggested that the learning mechanism involved computing the transitional probability of adjacent syllables, using the peaks in transitional probability to group syllables, into words, while using the dips in transitional probability to form breaks between words. Aslin, Saffran, and Newport (1998) provided evidence that the computation did indeed involve transitional probability rather than simple bigram or trigram frequency.

In this article, we extend these two lines of research and explore whether learners can use transitional probability as a cue to phrase structure. Our hypothesis is that they will be able to do so, and that grammars with rich transitional probability cues to phrase structure will be learned better than carefully matched control languages that lack such transitional probability cues.

TRANSITIONAL PROBABILITY

Transitional probability is a conditional probability statistic that measures the predictiveness of adjacent elements. By definition, the forward transitional probability[1] of successive elements XY is:

[1]There are also other conditionalized statistics, such as backward transitional probability, mutual information, conditional entropy, and correlation. All of these statistics are functionally equivalent in the research conducted thus far (including the research presented in this article), in that they normalize the frequency of co-occurrence of two elements by the frequency of one or both of those elements individually. Aslin et al. (1998) demonstrated that one of this class of conditionalized statistics predicts

$$\text{Probability of } Y | X = \frac{\text{frequency of } XY}{\text{frequency of } X}$$

Our hypothesis is that transitional probability computations facilitate the learning of linguistic structure by marking natural groupings—constituents, or phrases—in an otherwise uninterrupted stream of words. Saffran, Aslin, and Newport (1996) showed that learners can use transitional probability patterns between *syllables* to find *word* boundaries. We hypothesize that this procedure will work at a higher level of structure as well, and that learners will be able to use transitional probability patterns between *words* (or word classes) to find *phrase* boundaries. As Morgan et al. (1987, 1989) have shown, knowing where the phrases are in a stream of words greatly facilitates the acquisition of syntax, because many important syntactic dependencies occur within phrases (or between phrases themselves).

To determine whether learners could have used transitional probability as a cue to phrase boundaries in previous studies, we calculated the forward and backward transitional probabilities between classes of words in the grammars used by Moeser and Bregman (1972), Morgan and Newport (1981), and Morgan et al. (1987, 1989). It turned out that transitional probability (forward or backward) was not a cue that learners could have used to segment the speech stream into phrases in the languages they used. The transitional probabilities in these languages were no higher within phrases than at the boundaries of phrases. In retrospect, we realized that this might have been the reason that the phrase structure alone was not sufficient for learners to acquire these complex grammars and why additional cues to phrase boundaries (e.g., prosody) were necessary.

We also analyzed the forward and backward transitional probabilities between classes of words in the miniature grammar used by Saffran (2001). Saffran hypothesized that learning would be largely influenced by *perfect dependencies* between classes of words in this language. These dependencies were backward transitional probabilities of 1.0 between form classes in two of the three phrases. However, the forward transitional probabilities within those same phrases were relatively low. Overall, neither the backward nor forward transitional probabilities provided consistently good cues to phrase structure in the version of the grammar used by Saffran (2001). Although participants in Saffran's study did learn without the aid of additional cues such as prosody, they did not achieve the thorough learning of the grammar shown in Morgan et al.'s studies (1987, 1989), when participants had the advantage of such additional cues.

---

learning. *Transitional probability* was first used for psycholinguistic materials by Miller and Selfridge (1950), and, for the sake of simplicity, we refer to this statistic throughout the article. But it is important to note that our findings are also compatible with the claim that learners are computing another closely related statistic, such as *mutual information* or *conditional entropy.* We discuss these issues further in the General Discussion.

To explore the effect of transitional probability as a cue to phrase boundaries, we considered which distributional phenomena in a language would create peaks in transitional probability within phrases and dips in transitional probability at phrase boundaries. There are several features that a language could have that would create this pattern of transitional probability across a corpus of sentences. The experiments presented in this article investigate the effect of many such features, both when presented singly and in combination. Each of these features will be described in detail in the experiments below. But here we present an initial illustration of how a simple feature of a grammatical system can result in peaks in transitional probability within phrases and dips in transitional probability at phrase boundaries.

OPTIONAL PHRASES

Over a corpus of sentences, an interesting transitional probability pattern is created when words within phrases always remain together, but the phrases themselves may be optional or deleted from a sentence. Take as a starting point the simple example where a language has only six word classes (such as *noun* and *verb*). Representing each of these word classes with a single letter, A through F, in a very simple language a grammatical sentence might take the form ABCDEF, with the words subgrouped into the phrases AB, CD, and EF. Suppose the learner hears 10 sentences of this form, ABCDEF.

With only the sentence type ABCDEF in the language, there are no transitional probability patterns within sentences and no distributional evidence for phrases. A is always followed by B, and the transitional probability from A to B (within a phrase) is 1.0. Similarly, because B is always followed by C, the transitional probability from B to C (spanning a phrase boundary) is also 1.0. When all sentences take the same form, then, there is no transitional probability information for phrases within the sentence.

Suppose, however, that the CD phrase is optional, so that when the learner hears 10 sentences, 5 are of the form ABCDEF and 5 are of the form ABEF. The transitional probability from A to B and from C to D (within phrases) is still 1.0. However, now the transitional probability from B to C (spanning a phrase boundary) is only 0.5, because in half of the sentences B is now followed by E rather than C. In short, having an optional CD phrase, but always keeping A together with B and C together with D, maintains transitional probability peaks within phrases but creates a transitional probability dip at the boundary between phrases. This is just one example of how a common syntactic feature of natural languages, optional phrases, creates a pattern of transitional probabilities that could be used by a learner to induce phrase structure.

THE NATURAL LANGUAGE ENVIRONMENT

All of the features that we have included in our experiments are widespread in natural languages. For example, in English, prepositional phrases can be optional. One can say *The box on the counter is red* or *The box is red*, omitting the prepositional phrase *on the counter.*

We have included three other features in our experiments: repeated phrases, moved phrases, and different-sized form classes within phrases. Each of these is common in natural languages. Sentences often have more than one instance of the same type of phrase; for example, simple transitive sentences have two noun phrases, one in the subject position and one in the object position. Phrases may appear in moved or permuted order, as with passive or topicalized sentences. Finally, there are large differences among form classes in the number of words belonging to each class. Compare, for example, the size in English of the class of determiners (perhaps two dozen) to the class of nouns (tens of thousands). That these various syntactic features of language serve to demarcate phrases is not a recent discovery. In fact, these are precisely the features that linguists have used as distributional evidence for the *existence* of the phrase as a unit of language (Radford, 1988).

Morgan et al. (1987) discuss the role that the above features have in cueing the existence of phrases during learning. Their miniature artificial language contained most of these features (an optional phrase, a repeated phrase, and variation in the number of words assigned to different form classes). However, other features of the language, such as optional elements within each phrase and the fixed ordering of phrases, conspired to eliminate the transitional probability peaks within phrases and the dips at phrase boundaries. At that time, the authors were not thinking of transitional probability as the computational mechanism by which phrase structure might be acquired, and therefore did not consider the mathematical result of these additional syntactic features. The contribution of the present work is to suggest that these distributional phenomena may assist learning via a statistical route, analogous to the statistical learning mechanism previously shown for word segmentation.

THE PRESENT EXPERIMENTS

The aim of the present experiments, then, is to explore the effect that these syntactic features have, alone and in combination, on the transitional probability patterns between words, and to test whether learners can use these statistical patterns to learn the phrases and overall structure of a language. In the miniature languages that we have designed for the present studies, we have simplified the application of syntactic features (e.g., by making every phrase of a sentence optional with equal frequency); this is not meant to perfectly mimic the application of these features in natural languages, but rather to create a controlled statistical pattern that we can use to examine learners' sensitivities in a laboratory setting.

EXPERIMENT 1

Experiment 1 explored the effect of the syntactic feature discussed above—optional phrases—when applied to a miniature artificial language.

Method

Participants

Thirty-two monolingual English-speaking undergraduate students were recruited from the University of Rochester to participate in this study. All participants gave informed consent before participating. Participants received monetary compensation each day for 5 days, as well as a financial bonus on completion of the experiment.

Description of the Linguistic Systems

The two languages in this experiment—the optional phrases language and the optional control language—are both variations on a simple miniature language that we call the *baseline language*.

Baseline language. The baseline language has a simple linear structure, similar to the hypothetical language mentioned in the introduction. There is only one sentence type: ABCDEF. Each letter, A through F, represents a form class analogous to the lexical classes *noun* and *verb*. Three words are assigned to each form class. The 18 monosyllabic consonant–vowel–consonant (CVC) nonsense words of the language are listed in Table 1.[2]

Optional phrases. The optional phrases language uses the baseline language as a starting point, but in this language the six form classes are grouped in pairs (hereafter, *phrases*)[3] as follows: AB, CD, and EF. A grammatical sentence may consist of all three of these phrases, in that order, resulting in the sentence type ABCDEF (the

---

[2]The 18 CVCs (hereafter, *words*) were selected based on two criteria. First, each word scored between 70 and 80 on an index of the meaningfulness (to English speakers) of all possible CVC trigrams (Archer, 1960). Second, with respect to the neighborhood density of their phonotactics, none of the words appeared on the lists of either low-probability or high-probability nonwords in Vitevitch and Luce (1999). Hence, each CVC had a medium-probability neighborhood density of phonotactics.

[3]The only cue to this phrasal grouping is the pattern in the order of the words and the statistics created by this pattern. There are no other cues to phrase structure, such as pauses between phrases or intonation variations.

TABLE 1
Nonsense Words Assigned to Each Form Class

A Words	*B Words*	*C Words*	*D Words*	*E Words*	*F Words*
KOF (oaf)	HOX (box)	JES (dress)	SOT (coat)	FAL (pal)	KER (her)
DAZ (has)	NEB (web)	REL (fell)	ZOR (core)	TAF (waif)	NAV (have)
MER (her)	LEV (rev)	TID (bid)	LUM (bum)	RUD (bud)	SIB (bib)

Note. English pronunciation rhyme in parentheses.

canonical sentence type from the baseline language). Additionally, a grammatical sentence may be created by removing one of the three phrases and making it "optional." This results in a total of four distinct sentence types: ABCDEF, plus ABCD, ABEF, and CDEF. The optional phrases language can be represented by phrase structure rules: $S \rightarrow (P1) + (P2) + (P3); P1 \rightarrow A + B; P2 \rightarrow C + D; P3 \rightarrow E + F$, with the stipulation that every sentence must have at least two phrases. Example sentences are: MER HOX JES LUM TAF KER (a sentence with the structure ABCDEF) and KOF LEV RUD SIB (a sentence with the structure ABEF).

As we illustrated above, allowing phrases to be optional creates a pattern of transitional probability peaks within phrases and dips at phrase boundaries. Within each phrase, transitional probabilities are perfect—in other words, every A word is followed by a B word, every C word is followed by a D word, and every E word is followed by an F word. This maintains the transitional probability peak of 1.0 within phrases. However, a B word may now be followed by either a C word or an E word, creating a dip in transitional probability at the phrase boundary. Similarly, a D word can now be followed by an E word, or it can be the end of the sentence, creating a dip in transitional probability at the boundary between D and E. The transitional probability information that participants in this condition were exposed to is shown in Table 2 (along with the transitional probability patterns for the languages of the other experiments in this article).[4]

Optional control. The critical feature for the optional control language is that it lacks the pattern of peaks and dips in transitional probability between word classes. The baseline language described previously fits this criterion and could be considered a control condition for the optional phrases language. However, the optional phrases language is quite different from the baseline language: it has sentences that are both four and six words long, and the form classes do not occupy

[4]Table 2 shows the major transitional probabilities for each condition—that is, the transitional probabilities that characterize the baseline sentences for each language. Of course, participants are exposed to other nonzero transitional probabilities created by the specific manipulations of each condition, but these are always relatively small. For example, the transition from B to C (as shown in Table 2) is 0.8, but B can also be followed by E (in sentences where the CD phrase is optional), with a transitional probability of 0.2.

TABLE 2
Transitional Probabilities From Experiments 1 Through 4

	$A \to B$	$B \to C$	$C \to D$	$D \to E$	$E \to F$
Optional phrases	1.00	0.80	1.00	0.80	1.00
Optional control	0.90	0.90	0.90	0.90	0.90
Repeated phrases	1.00	0.86	1.00	0.86	1.00
Repeated control	0.92	0.94	0.92	0.94	0.93
Moved phrases	1.00	0.60	1.00	0.60	1.00
Moved control	0.78	0.78	0.78	0.78	0.78
Class size variation	1.00	1.00	1.00	1.00	1.00
Class size control	1.00	1.00	1.00	1.00	1.00
All-combined	1.00	0.56	1.00	0.52	1.00
All-combined control	0.80	0.83	0.74	0.77	0.74
All-combined 5%	1.00	0.33	1.00	0.22	1.00
All-combined 5% control	0.67	0.71	0.58	0.59	0.47

fixed positions within every sentence. To control for these complexities, we designed a control language that matched the optional phrases language in as many ways as possible.

In the optional control language, there are still six form classes, A through F, with the same words assigned to these classes as in the optional language (see Table 1). Also, in addition to the canonical ABCDEF sentence type, there are four-word sentences, created by removing adjacent pairs of words. Hence each form class no longer occupies a fixed position in every sentence. The critical difference between the two languages is that in this language, the optional word pairs can be any adjacent pair of words: any one of the pairs AB, BC, CD, DE, EF, or AF can be optional in a sentence.[5] Example sentences are: KOF NEB REL SOT FAL NAV (a sentence with the structure ABCDEF) and LEV TID ZOR RUD (a sentence with the structure BCDE).

Notice that what this does to the statistics of the optional control language. Because A is no longer always followed by B, the transitional probability of AB is no longer 1.0. In the same fashion, the 1.0 transitional probability within each of the other phrases has been reduced. In addition, transitional probabilities between phrases have been increased. The result is a pattern of flat transitional probability statistics, with no variation—no peaks and dips—between form classes. The transitional probability information that participants in this condition were exposed to is shown in Table 2. It is important to note that the optional control language cannot be written with phrase structure rules, as it is not a phrase structure grammar (this will be true as well of the control languages in all subsequent experiments).

---

[5]We allowed the AF pair to be optional, even though A and F are not adjacent. This was necessary for the endpoint classes to have statistical distributions similar to the other form classes.

Materials

The optional phrases presentation set. The presentation set for this condition consisted of 96 of the possible 972 grammatical sentences of the language. To reflect the fact that natural languages often have a canonical sentence type that occurs more frequently than derived sentence types, half of these, or 48 sentences, were of the canonical (ABCDEF) form and were selected randomly from all possible sentences of that type. For the remaining 48 sentences, 16 sentences were selected randomly from each of the ABCD, CDEF, and ABEF types. All 96 sentences were randomized and concatenated to form the presentation set.

The optional control presentation set. The presentation set for the control condition was constructed in a similar way. Half of the sentences, or 48 sentences, had the structure ABCDEF and were identical, in type and token, to the 48 ABCDEF sentences in the experimental condition. The rest of the sentences were only four words long, one-sixth from each of the six possible sentence types that had a pair of words omitted. Half of these sentences were also taken directly from the experimental condition (the ones that were also legal sentence types in that condition). The remaining 24 sentences, of types ADEF, ABCF, and BCDE, were unique to the optional control condition. All 96 sentences were randomized and concatenated to form the presentation set.

Recording. Each sentence was individually recorded into a Dell Pentium II PC using a Kay CSL 4300B sound editor with a head-mounted AKG C 420 MicroMic II microphone. Sentences were carefully spoken with uniformly descending prosody (list intonation) by a trained female speaker. To ensure consistent speed, word tempo was standardized using a metronome set to 195 beats per minute. A random sample of sentences was selected for detailed acoustical analysis. The analysis revealed no significant or systematic differences (in pitch contour, stress, vowel duration, pauses between words, etc.) between words within phrases versus words spanning phrase boundaries.

Within the Kay CSL sound editor, the sentences for each presentation set were concatenated in random order, with a 1.4-s intersentence interval (isi). Each presentation set was then transferred into SoundEdit 16 (Version 2) and concatenated four times to form an input soundfile of approximately 20 min duration.

Procedure

The experiment was administered individually via a Psyscope 1.2.5 PPC program inside a small private room with an iMac and Sennheiser HD 570 headphones. Participants were told to "have a seat, wear the headphones, and follow the instructions on the screen—the experiment will be self-explanatory."

The complete procedure, administered on Days 1 and 5 of the experiment, consisted of a "pretest" to familiarize subjects with button pressing in preparation for the later tests, 20 min of exposure to the language, and two tests covering what they learned (see below). On Days 2, 3, and 4, participants just listened to the language. Because 20 min of exposure to the language, consisting of four repetitions of the presentation set, was given on each of the five consecutive days, at the end of the fifth day each participant had heard the presentation set 20 times.

Sentence Test

The Sentence Test was a forced-choice test designed to test participants' knowledge of the linear word order rules of the language. It tested only the canonical sentence type, ABCDEF. There were 30 items on the Sentence Test, each consisting of a pair of sentences. One sentence was a novel canonical sentence (ABCDEF) that did not appear in either of the presentation sets. The other sentence was identical to the first, except that one word was removed and replaced by a word from another form class (with the constraint that the replacement word did not already appear in the sentence). Hence, the A word could be removed and replaced with a B word. Or the C word could be removed and replaced with an F word, and so on. There were six possible form classes that could be removed and five possible form classes that could replace it, resulting in the 30 different types of wrong answers for the test.

The sentences for the Sentence Test were recorded in list intonation by the same trained female speaker that recorded the sentences for the presentation sets, but at a slightly slower rate of 175 beats per minute. Each trial consisted of a grammatical sentence and its slightly altered counterpart with 1 s of silence between them. The right and wrong answers were the first member of the pair equally often, in random order throughout the test. The 30 trials appeared in the same randomized sequence for all participants.

Phrase Test

The Phrase Test was designed to assess the extent to which participants grouped the words together as phrases in their mental representation of the language. The test was forced-choice and consisted of 18 items, 6 items testing each of the three phrase types (AB, CD, EF). An item consisted of two pairs of words, one pair that constituted a phrase (e.g., AB) and one pair that was a legal sequence in the language but spanned a phrase boundary (e.g., BC). The phrasal pair was considered to be the right answer, whereas the pair that spanned a phrase boundary was considered to be the wrong answer (even though it was a grammatical sequence of words). Of course, for the control condition, the terms *right* and *wrong* answer are misnomers: there is no reason to expect that control subjects will prefer any pair of words over another. But we scored control subjects on the phrase test to establish a baseline of performance against which to compare performance in the experimental condition.

Each pair of words was separately recorded by the same female speaker that recorded the sentences for the presentation sets, but at a slightly slower pace of 175 beats per minute. The pairs were presented to the participant with 1 s of silence between them. The right answer was the first pair or the second pair equally often, in random order throughout the test. The 18 trials appeared in the same randomized sequence for all participants. Before the test began and as a prompt after each trial, participants received instructions that they should choose the pair of words that sounded "more like a group or unit from the language."

Results

Sentence Test

Our first question was whether participants in the experimental condition learned the basic structure of the language (e.g., the canonical sentence type—ABCDEF) better than controls. To answer this question, we analyzed performance on the Sentence Test with one-tailed *t* tests. Figure 1 shows the results of the Sentence Test on Days 1 and 5. On Day 1, the optional phrases condition outperformed the optional control condition, $t(1, 30) = -1.69$, $p = .05$. This difference remained on Day 5, $t(1, 30) = -1.99$, $p = .028$, indicating that participants in the optional phrases condition did indeed learn the basic word order of the language significantly better than did participants in the optional control condition. As described

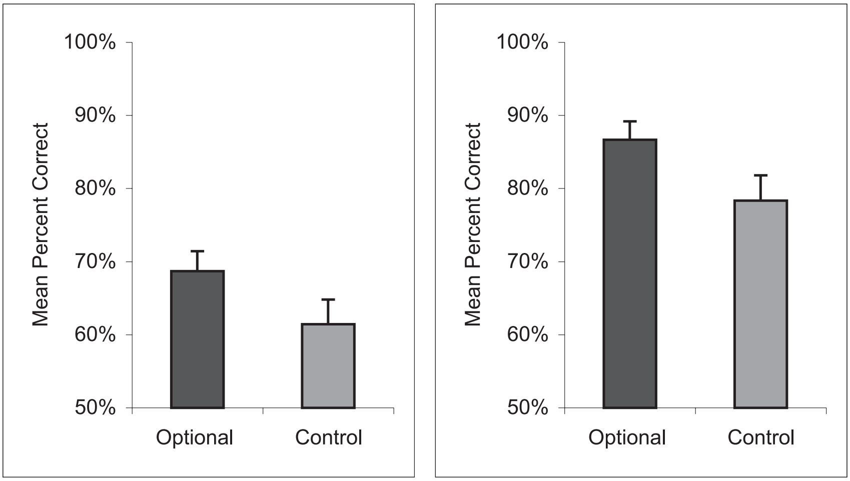


FIGURE 1 Experiment 1: Results of the Sentence Test on Day 1 (left) and Day 5 (right).

above, the Sentence Test contained all novel sentences that were not in the presentation set. These results thus indicate that participants acquired the ordering of word classes and did not merely remember specific sentences they had been exposed to.

Phrase Test

The question that we were most interested in was whether participants in the optional phrases condition learned their language as organized in terms of phrase structure. We hypothesized that the pattern of peaks and dips in transitional probability present in the experimental condition would cause those participants to form a more hierarchically structured grammar than controls—a grammar with an added level of phrasal representation. To determine whether this was the case, we analyzed the results of the Phrase Test, shown in Figure 2. On Day 1, the optional phrases condition outperformed the optional control condition in a one-tailed t test: $t(1, 30) = -2.83$, $p = .004$. This highly significant difference remained on Day 5, $t(1, 30) = -3.16$, $p = .0018$, indicating that the participants in the optional phrases condition did indeed learn a hierarchical phrase structure of the language significantly better than did participants in the optional control condition.

Discussion

The results of the Sentence Test show a significant difference between the optional phrases and optional control conditions on both Days 1 and 5, with participants in

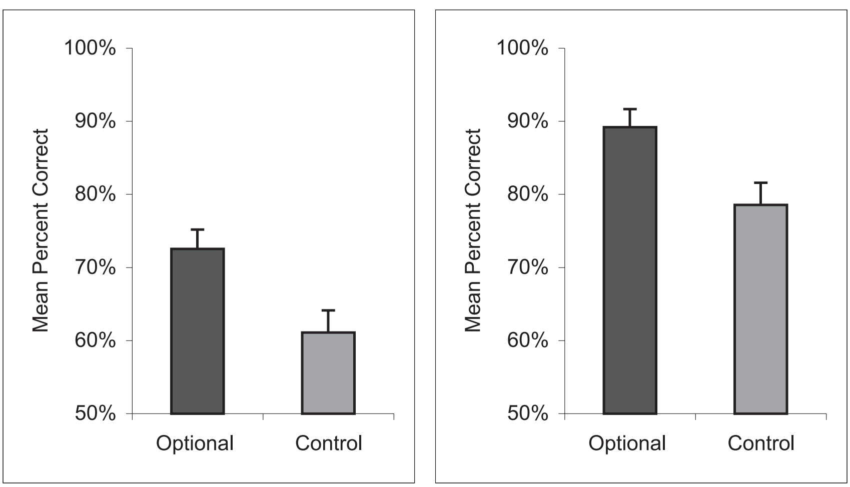


FIGURE 2 Experiment 1: Results of the Phrase Test on Day 1 (left) and Day 5 (right).

the experimental condition outperforming controls on a test of the basic word order of the language. Both conditions showed learning from Day 1 to Day 5. Despite the difference between the conditions, it is worth noting how well participants in the optional control condition did, particularly on Day 5. This is perhaps not surprising. The Sentence Test was a simple test of the linear structure of the canonical sentence type, a sentence structure to which participants in both conditions were heavily exposed. Moreover, participants could score well on the Sentence Test by simply learning the correct position for each word. Indeed, it is perhaps remarkable that any difference emerged between the two groups, given the simplicity of the test and the amount of exposure to the canonical sentence type (a total of 960 sentence tokens) that participants in each group received.

Our hypothesis was that the difference between the groups emerged as a result of the different type of language structure induced by participants in each group. Surprisingly, participants in the control condition scored well above chance on the Phrase Test; we return to consider some reasons for this in the next experiment. Importantly, however, the results on the Phrase Test indicate that participants in the optional phrases condition formed a significantly stronger representation of phrases than did participants in the optional control condition. This difference was highly significant on Days 1 and 5. Because participants in the optional control condition formed a weaker representation of phrases, they were more likely to make judgments on the Sentence Test based solely on each word's position within a sentence string. Participants in the optional phrases condition, however, formed a strongly hierarchical representation of the language, with words grouped together as phrases and these phrases strung together into sentences. Because of this added layer of representation, these participants were more likely to make judgments *both* based on each word's legal position in a sentence *and* based on their knowledge of the proper pairing of words into phrases.[6]

The results of Experiment 1 demonstrate that a simple feature of natural languages, optional phrases, when added to a miniature artificial language, creates a pattern of transitional probability peaks within phrases and transitional probability dips at phrase boundaries; that learners are sensitive to this statistical pattern; and that they do in fact use it to learn fully the hierarchical phrase structure and overall grammar of the language. The next step is to test other syntactic features common in natural languages that can provide distributional evidence for phrase structure. We do this in Experiment 2.

[6]Several reviewers have asked whether our results could arise instead from serial position effects; that is, could better learning of beginnings and ends of sentences (sometimes seen in other miniature language experiments; see Morgan et al., 1987) interact with our syntactic manipulations to produce the overall better learning in the experimental condition? We discuss this hypothesis more fully in the General Discussion section, but for now it is important to note that there are no significant differences in any of our experiments between learning of the beginnings and ends of sentences and those of sentence middles, and no significant interactions between this serial position contrast and condition.

EXPERIMENT 2

In Experiment 2 we expand the paradigm established in Experiment 1, testing three additional features of natural languages: repeated phrases, moved phrases, and variation in the size of word classes.

Each of these features serves to create peaks in transitional probability between words within phrases, and dips in transitional probability between words at phrase boundaries. Therefore we would expect participants in all three experimental conditions to perform similarly to participants in the optional phrases condition. This was our general hypothesis. However, there is one caveat to this expectation. Although the repeated-phrases and moved-phrases features create peaks and dips in transitional probability in a manner similar to the optional phrases feature, the class size variation feature creates a somewhat different statistical pattern.

With optional, repeated, and moved phrases, the pattern of peaks and dips in transitional probability occurs *both* at the level of word classes *and* between the individual words themselves. In contrast, with class size variation, there is *no* pattern of peaks and dips at the level of word *classes*. The pattern of peaks and dips occurs only between the individual words themselves. Thus, if participants are (implicitly) tabulating the transitional probabilities between individual words, then the results of the class size variation condition should resemble the results from the other conditions. If, however, participants are forming classes out of the words and tabulating the transitional probabilities between these word classes, then the pattern of results in the class size variation condition should deviate from the other conditions. Specifically, participants should fail to form a hierarchical phrase structure representation of the language, and should not score significantly higher on the Phrase Test than controls.

Method

Participants

A total of 98 monolingual English-speaking undergraduate students were recruited from the University of Rochester to participate in this study. All participants gave informed consent and were paid for their participation.

Description of the Linguistic Systems

Repeated phrases. Like the optional phrases language from Experiment 1, the repeated phrases language uses the baseline language as a starting point, and goes on to group the six form classes together into the phrases AB, CD, and EF. In the repeated phrases language, a grammatical sentence may be either the canonical sentence type, or ABCDEF plus one phrase repeated at the end. Thus there are four distinct sentence types: ABCDEF, as well as ABCDEFAB, ABCDEFCD, and

ABCDEFEF. No individual word may be repeated in a sentence. The repeated phrases language can be represented by phrase structure rules as follows: S → P1 + P2 + P3 + (*P); P1 → A + B; P2 → C + D; P3 → E + F; *P → {P1, P2, P3}. Example sentences are: DAZ LEV JES SOT TAF SIB (a sentence with the structure ABCDEF) and MER HOX JES ZOR FAL NAV TID SOT (a sentence with the structure ABCDEFCD).

Over a corpus of sentences, allowing phrases to be repeated creates a pattern of transitional probability peaks within phrases and transitional probability dips at phrase boundaries. Within each phrase, transitional probabilities are still perfect, because A is still always followed by B, C is still always followed by D, and so on. However, B can now either be followed by C or by the end of the sentence. D also can either be followed by E or by the end of the sentence. And F can now be followed by the end of the sentence or by A, C, or E. The transitional probability pattern that participants in this condition were exposed to is shown in Table 2.

The presentation set for this condition consisted of 68 grammatical sentences.[7] Half of these, or 34 sentences, were of the canonical sentence type. The other half were divided (as equally as possible) between the other three sentence types (ABCDEFAB, ABCDEFCD, and ABCDEFEF), with specific sentences chosen randomly from all possible sentences of that type. All sentences in Experiment 2 were recorded as per the procedure described in Experiment 1. The 68 sentences were randomized and concatenated, with a 1.8-s isi, to form the presentation set.

Repeated control. In like fashion to the optional control language, we constructed a repeated control language without the critical peaks and dips in transitional probability, but that matched the repeated phrases language as closely as possible in all other ways. In the repeated control language, a sentence may either be canonical, or may be formed by adding any of the pairs AB, BC, CD, DE, EF, or AF to the end of a canonical sentence. The effect of this manipulation on the transitional probability pattern that participants were exposed to is shown in Table 2. Example sentences are: MER NEB TID ZOR TAF SIB (a sentence with the structure ABCDEF) and DAZ LEV JES SOT RUD NAV HOX REL (a sentence with the structure ABCDEFBC).

The presentation set for this condition had 68 sentences, of which 51 were identical (in type and token) with sentences from the repeated phrases condition (this included all sentences of types ABCDEF, ABCDEFAB, ABCDEFCD, and ABCDEFEF). The remaining 17 sentences were unique to the repeated control

[7]Across experiments and conditions, the different syntactic features (e.g., optional versus repeated phrases) created different length sentences. We therefore equated conditions for the total number of words (and phrases) in the presentation sets rather than for the number of sentences. We then adjusted the isi slightly to equate conditions for overall time of exposure.

condition. All 68 sentences were randomized and concatenated, with a 1.8-s isi, to form the presentation set.

Moved phrases. The moved phrases language uses the baseline language as a starting point, and goes on to group the six form classes together into the phrases AB, CD, and EF. In this language, a grammatical sentence may be the canonical sentence type, ABCDEF, or a sentence with any of the three phrases in moved positions. Hence ABCDEF, ABEFCD, CDABEF, CDEFAB, EFABCD, and EFCDAB are the six legal sentence types in the language, created by moving (permuting) the three phrases in all possible ways. The moved phrases language can be represented by phrase structure rules as follows: S $\rightarrow$ P1 + P2 + P3; P1 $\rightarrow$ A + B; P2 $\rightarrow$ C + D; P3 $\rightarrow$ E + F, with a movement rule that reads *Move any phrase anywhere*. Example sentences are: DAZ NEB JES LUM TAF NAV (a sentence with the structure ABCDEF) and RUD SIB TID LUM MER LEV (a sentence with the structure EFCDAB).

Over a corpus of sentences, allowing phrases to be moved creates a pattern of transitional probability peaks within phrases and transitional probability dips at phrase boundaries. Within each phrase, transitional probabilities are still perfect, because A is still always followed by B, C is still always followed by D, and so on. However, B can now be followed by C, E, or the end of the sentence. D can also be followed by E, A, or the end of the sentence. And F can now be followed by the end of the sentence or by A or C. The transitional probability pattern that participants in this condition were exposed to is shown in Table 2.

The presentation set for this condition consisted of 80 grammatical sentences, half of which were of the canonical sentence type. The remaining 40 sentences were divided equally between the other five sentence types, with specific sentences chosen randomly from all possible sentences of a particular type. The 80 sentences were randomized and concatenated with a 1.6-s isi to form the presentation set.

Moved control. In like fashion to the previous pairs of conditions, we constructed a matched control for the moved phrases language. This control language has moved pairs of words, but does not have the critical pattern of peaks and dips in transitional probability structure. To flatten the transitional probabilities between words (and word classes) in this language, we added four more sentence types: BCAFDE, AFDEBC, DEAFBC, and DEBCAF. These sentence types were created by moving word pairs that are not phrases, namely, BC, DE, and AF. The transitional probability pattern that participants in this condition were exposed to is shown in Table 2. Example sentences are: MER LEV JES SOT FAL KER (a sentence with the structure ABCDEF) and LEV TID MER SIB SOT TAF (a sentence with the structure BCAFDE).

For the presentation set, half of the 80 sentences had the canonical sentence structure and were identical in type and token to the canonical sentences used in

TABLE 3
Word Categories From the Class Size Variation Language

A words	*B words*	*C words*	*D words*	*E words*	*F words*
KOF (oaf)		JES (dress)		FAL (pal)	
DAZ (has)	NEB (web)	REL (fell)	ZOR (core)	TAF (waif)	NAV (have)
MER (her)	LEV (rev)	TID (bid)	LUM (bum)	RUD (bud)	SIB (bib)
HOX (box)		SOT (coat)		KER (her)	

Note. English pronunciation rhyme in parentheses.

the moved phrases condition. The other half of the sentences were "moved" versions. Of these, 22 were identical in type and token to sentences in the moved phrases condition, and 18 were unique to the control condition. All 80 sentences were randomized and concatenated with a 1.6-s isi to form the presentation set.

Class size variation. The class size variation language has a simple linear structure (it is not a phrase structure grammar), with only one sentence type, the canonical sentence ABCDEF. In fact, the only difference between this language and the baseline language is the assignment of words to form classes: classes A, C, and E have four possible words, whereas the classes B, D, and F have two possible words. This new word assignment is shown in Table 3. An example sentence is: HOX LEV SOT LUM KER SIB (a sentence with the structure ABCDEF).

Because there is only one sentence type, and because every A word is followed by a B word, every B word by a C word, and so on, throughout the sentence, the transitional probability between word classes is 1.0 across the whole sentence, as shown in Table 2. However, at the level of individual words, there is variation in transitional probability. Each A word has two possible B words that could follow it, so each A-to-B word transition has a .5 transitional probability. Each B word, however, could be followed by any one of four C words, so each of these transitions has a .25 transitional probability. The transitional probabilities between individual words in this language are shown in Table 4.[8]

The presentation set for this condition was created by randomly selecting 80 sentences from the language and concatenating them with a 1.6-s isi.

Class size control. The control language for the class size variation condition is simply the baseline language that was described in Experiment 1. The varia-

[8]This variation in the number of words per class creates classes of words that are high in frequency, namely, the B, D, and F classes. (Because there are only two words in each of these classes, these words will be relatively frequent in the input set.) This may facilitate learning (for evidence that having high frequency markers in a miniature language facilitates grammar learning, see Valian & Coulson, 1988).

TABLE 4
Transitional Probabilities Between Individual Words in the Class Size Variation and Class Size Control Languages

	DAZ→NEB	*NEB→REL*	*REL→ZOR*	*ZOR→TAF*	*TAF→NAV*
Class size variation	.50	.25	.50	.25	.50
Class size control	.33	.33	.33	.33	.33

tion in transitional probability at the level of individual words is removed by going back to the word assignments shown in Table 1. Hence, this control language has no variation in transitional probability either between form classes, as shown in Table 2, or between individual words, as shown in Table 4. An example sentence is: KOF HOX JES SOT FAL KER (a sentence with the structure ABCDEF). The presentation set was created by randomly selecting 80 sentences from the baseline language and concatenating them with a 1.6-s isi.

Procedure

The procedure for Experiment 2 was identical to the procedure for Experiment 1. Participants in all conditions were exposed to their respective presentation sets four times, for a total of about 20 min of exposure, on each of 5 consecutive days. On Days 1 and 5 of the experiment, after being exposed to the language, participants in all conditions received the same Sentence Test and Phrase Test from Experiment 1.[9]

Results

Sentence Test

Our first question was whether participants in the experimental conditions learned the basic word order of the canonical sentence type better than the controls. To answer this question, we analyzed the results of the Sentence Test. The results from the Sentence Test on Day 1 and Day 5 are shown in Figure 3, along with the analogous results from Experiment 1, for comparison.

We performed an analysis of variance with Condition (moved, repeated, class size) and Treatment (experimental versus control) as the two between-subjects factors. On both Day 1 and Day 5 there was a significant main effect of treat-

[9]Even though the class size variation condition had a different assignment of words to word classes, we were able to use the same Sentence Test because it used only those items (15/18 words) that had identical assignments in all languages. The Phrase Test, however, had to be slightly altered for the class size condition (this modified Phrase Test was also used for the all-combined conditions in Experiments 3 and 4).

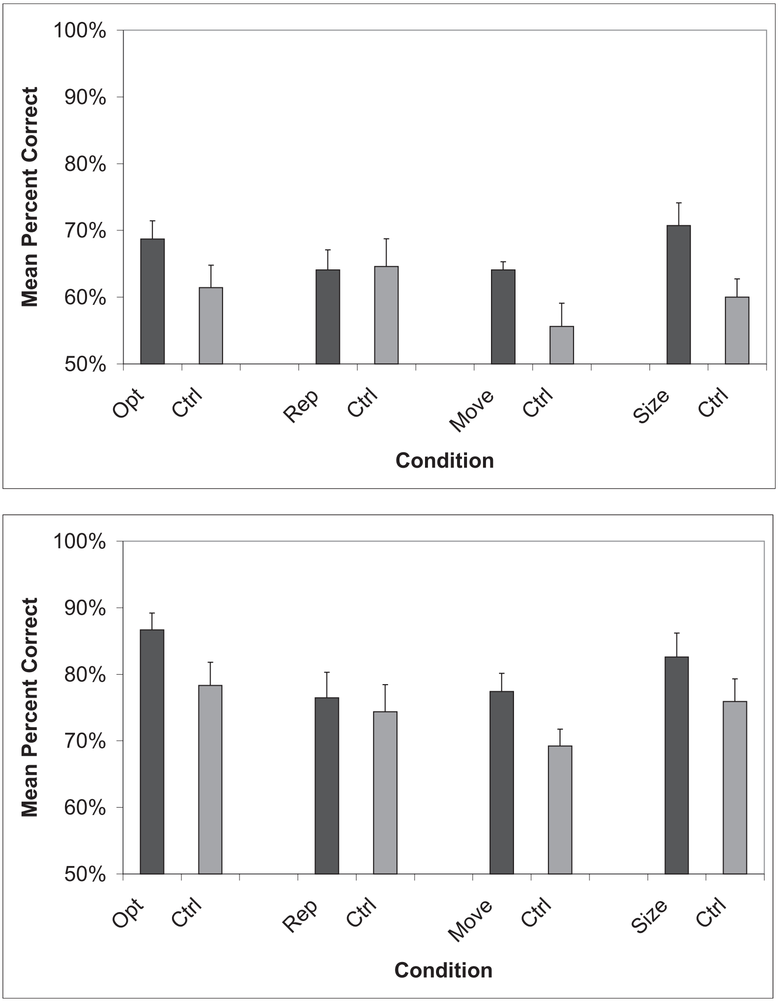


FIGURE 3 Experiment 2: Results of the Sentence Test on Day 1 (top panel) and Day 5 (bottom panel).

ment: for Day 1, $F(1, 92) = 5.85$, $p = .018$; for Day 5, $F(1, 92) = 3.96$, $p = .049$. There was no main effect of condition on either day—for Day 1, $F(2, 92) = 1.73$, $p = .18$, *ns*; for Day 5, $F(2, 92) = 1.57$, $p = .21$, *ns*—and no treatment by condition interaction: for Day 1, $F(2, 92) = 1.79$, $p = .17$, *ns*; for Day 5, $F(2, 92) = .39$, $p = .68$, *ns*. These results indicate that participants in the experimental groups did learn the basic word order of the language better than participants in control groups.

Phrase Test

The question that we were most interested in was whether participants in the experimental conditions formed a hierarchical phrase structure representation of the language. A secondary question of interest was whether the class size variation pair of conditions performed similarly to the other pairs of conditions. To answer these questions, we analyzed the results of the Phrase Test, which are shown in Figure 4.

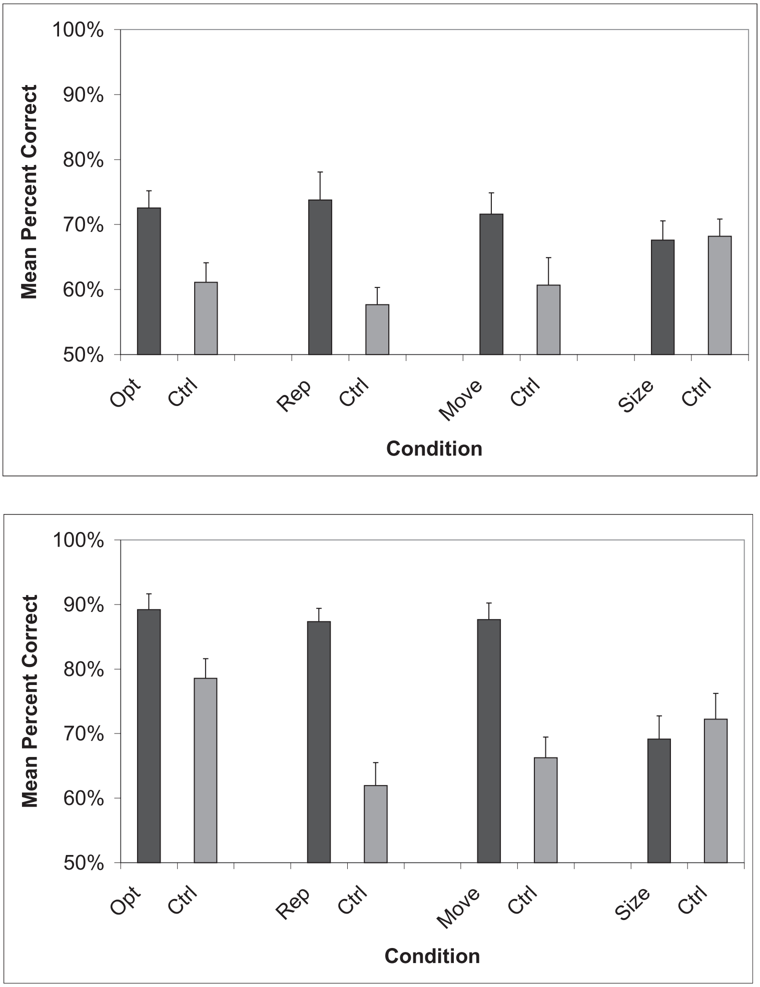


FIGURE 4 Experiment 2: Results of the Phrase Test on Day 1 (top panel) and Day 5 (bottom panel).

Again we performed an analysis of variance with Condition (moved, repeated, class size) and Treatment (experimental versus control) as the two between-subjects factors. On Day 1, there was a significant main effect of treatment, $F(1, 92) = 9.58$, $p = .003$, no main effect of condition, $F(2, 92) = .23$, $p = .79$, *ns*, and a significant treatment by condition interaction, $F(2, 92) = 3.17$, $p = .047$, indicating that, across conditions, experimental groups outperformed controls on the phrase test, but that there was a difference in the extent to which experimental groups scored differently than control groups. This was because the repeated phrases and moved phrases conditions scored significantly higher than their matched control conditions—$F(1, 92) = 10.10$, $p = .002$ and $F(1, 92) = 4.66$, $p = .033$, respectively—whereas the class size variation condition did not outperform the class size control condition, $F(1, 92) = .02$, $p = .89$, *ns*.

A similar pattern of results emerged on Day 5. In the analysis of variance there was a highly significant main effect of treatment, $F(1, 92) = 29.47$, $p = .000$, no main effect of condition, $F(2, 92) = 1.95$, $p = .15$, *ns*, and a significant treatment by condition interaction, $F(2, 92) = 11.65$, $p = .000$. The interaction occurred because, again, the repeated phrases and moved phrases conditions scored significantly higher than their matched control conditions—$F(1, 92) = 28.21$, $p = .000$ and $F(1, 92) = 20.08$, $p = .000$, respectively—whereas the class size variation condition did not score differently from the class size control condition, $F(1, 92) = .50$, $p = .48$, *ns*.

Discussion

The results of the Sentence Test indicate that, overall, participants in the experimental conditions learned the basic word order of the language better than the controls. However, the results of the Sentence Test, particularly when compared with the results of the Phrase Test, were rather moderate. The structure of the input and the simplicity of the Sentence Test in our experiment—specifically, the fact that all participants heard 50% ABCDEF sentences (680–800 sentence tokens) and that success on the Sentence Test could be achieved by learning the relative position for each individual word—may have conspired to keep differences between experimental and control groups on the Sentence Test somewhat small. However, it is notable that despite these factors, a significant overall difference emerged between experimental and control groups.

The results on the Phrase Test were very strong, demonstrating that transitional probability is a powerful cue to phrase structure. The pattern of performance on the Phrase Test supports the hypothesis that adult learners can calculate the transitional probability of adjacent elements and use peaks in transitional probability to cue phrasal groupings and dips in transitional probability to signal the breaks between phrases.

Importantly, however, this pattern of results on the Phrase Test did not appear in the class size variation languages. In contrast to the other syntactic manipulations, the class size variation condition did not outscore its control condition on the Phrase Test, either on Day 1 or on Day 5. In addition, scores in these two conditions showed no improvement from Day 1 to Day 5. It was mentioned at the outset that these conditions tested a feature with statistical consequences fundamentally different from those of the other conditions. In the class size variation condition, dips in transitional probability at the hypothesized phrase boundaries occurred only in word-to-word transitions, but not in form class transitions.[10] The difference in the pattern of results for the class size variation conditions from that shown by each of the other pairs of conditions suggests that learners grouped words into word classes and were sensitive to the transitional probability peaks and dips between these word classes. Apparently, transitional probability patterns between individual words, which were equivalent for the class size conditions and the other experimental conditions, were not adequate to drive differential phrase structure learning. Taken together, these results provide evidence for the notion that learners in the other conditions were computing word class transitional probabilities.

One unexpected finding is that participants in all of the control conditions scored above chance on the Phrase Test, even though there were no cues to phrase structure in these conditions. Our main finding is that learners in the experimental conditions substantially outscored these control participants, showing clearly that transitional probability cues had a substantial effect on the learning of phrasal groupings. But why did control participants score above chance on this test?

Apparently, learners have a tendency to organize serially presented auditory input into binary groupings, even in the absence of other grouping information. Participants in these control conditions imposed an AB, CD, EF grouping pattern onto the input. Perhaps our native English-speaking participants were imposing a grouping structure similar to the rhythmic structure of English. Most English words have a trochaic structure, with a strong (stressed) syllable followed by a weak (unstressed) syllable (e.g., *happy*, *baby*, *sunshine*, and *teacher*). Participants might have imposed such a grouping as they listened to and stored input strings, even though no rhythmic information was physically present. Another possibility (in accord with X-bar theory; Jackendoff, 1977) is that learners might tend universally to organize word strings into binary phrasal groupings. A binary grouping hypothesis makes a further prediction about our data—that participants in the optional control condition would do better on the Phrase Test than participants in the moved control or repeated control conditions would. If participants adopted a binary grouping strategy, always breaking sentences into

[10]The other experimental conditions also had peaks and dips in transitional probability between individual words; however, these peaks and dips were quite similar to those between word classes. Specifically, they were each approximately one third of the transitional probabilities between word classes (because there were three words per class).

chunks, two words at a time, the resulting chunks might be right answers or wrong answers on the Phrase Test (or neither). As it turns out, only one of the chunks that is created when a nonphrasal word pair is *optional* is a wrong answer on the Phrase Test, whereas nearly all of the chunks that are created when nonphrasal word pairs are *repeated* or *moved* favor the wrong answers on the Phrase Test. Interestingly, the optional control condition did score higher on the Phrase Test than did the other control conditions. This provides additional evidence for the hypothesis that participants were naturally chunking the input two words at a time and that it was this tendency that resulted in above-chance learning of phrases where no phrases existed in the input sentences in the control conditions. Crucially, however, all the experimental conditions (except the class size variation condition) substantially exceeded the tendency of their control conditions to group words into phrases, due to the pattern of transitional probabilities they experienced.

EXPERIMENT 3

The results from Experiment 2 were promising, demonstrating that learners can use the transitional probability patterns created by many different syntactic features of natural languages to cue the existence of phrases, and thereby learn the constituent structure and word order of an artificial language. However, the languages were very simple, with only one phrasal syntactic feature introduced in each language. Natural languages, in contrast, have multiple syntactic features of this kind present in combination. It is not clear how participants would handle the complexity that would result if all of the features that we tested were present in combination. On the one hand, the resulting complexity could overload learners' computational abilities, resulting in little learning of hierarchical phrase structure. On the other hand, although presenting all four syntactic features in combination would result in a more complex language, it would also result in more pronounced transitional probability dips at phrase boundaries. Perhaps learners would respond favorably to this increase in structured complexity and use it to form even stronger hierarchical phrase structure representations of the language. To explore these possibilities, in Experiment 3 we tested a language, the "all-combined" language, that combined all four features that were present individually in Experiments 1 and 2.

Method

Participants

Twenty-four monolingual English-speaking undergraduate students were recruited from the University of Rochester to participate in this study. All participants gave informed consent and were paid for their participation.

Description of the Linguistic Systems

All-combined language. The language in the all-combined condition has all of the syntactic features, in one language, that were present individually in the previous experimental conditions. There are still six form classes, A through F. To add the feature of class size variation, the assignment of words to word classes is the same as in the class size variation condition (see Table 3). This creates the variation in transitional probability between individual words that was present in the class size variation condition. In addition, the six form classes are grouped together into the phrases AB, CD, and EF by all of the earlier phrasal manipulations. In other words, the phrases may be optional, repeated, in moved order, or in any combination of these manipulations, in one sentence. A legal sentence consists of anywhere between two and four phrases (four to eight words). A phrase may appear in a sentence no more than twice. No specific words may be repeated in a sentence. The all-combined language can be represented by phrase structure rules as follows: $S \rightarrow {*P} + {*P} + ({*P}) + ({*P})$; $P1 \rightarrow A + B$; $P2 \rightarrow C + D$; $P3 \rightarrow E + F$; ${*P} \rightarrow \{P1, P2, P3\}$. Example sentences are: FAL SIB TID ZOR (a sentence with the structure EFCD) and TID LUM TAF NAV KOF LEV JES ZOR (a sentence with the structure CDEFABCD).

Over a corpus of sentences, this combination of syntactic features creates a pattern of transitional probability peaks within phrases and transitional probability dips at phrase boundaries. Within each phrase, transitional probabilities remain a perfect 1.0. However, the dips in transitional probability at the boundaries of phrases are deeper than in any of the previous experimental conditions, as shown in Table 2.

The presentation set for this condition had 74 sentences, half of which were of the canonical sentence type ABCDEF. The other half of the sentences combined optional phrases, repeated phrases, and/or moved phrases in various ways, with one sentence typically exhibiting two or more of these features. All sentences in Experiment 3 were recorded per the procedure described in Experiment 1. The sentences were randomized and concatenated, with a 1.6-s isi, to form the presentation set, which was looped and played to participants for a total exposure time of approximately 20 min, on each of 5 consecutive days.

We have mentioned that, although the all-combined language has more transitional probability information (a slightly deeper dip at phrase boundaries) than any of the languages tested in Experiments 1 and 2, it is also a much more complex language. To illustrate this, Table 5 compares the experimental languages used so far in overall complexity, in terms of the number of sentence types and the total number of sentences in each language. By these metrics, the all-combined language should be much *harder to learn*. However, according to our hypothesis, the increase in transitional probability variation within versus between phrases should make this language *easier to learn*. In addition, the all-combined language pro-

TABLE 5
Complexity Comparison of the All-Combined Language and the Experimental Languages From Experiments 1 and 2

Language	*Sentence Types*	*Sentences*
Optional phrases	4	972
Repeated phrases	4	20,412
Moved phrases	6	4,374
Class size variation	1	512
All-combined	86	233,536

vides *correlated cues* to phrase structure, analogous to the correlated cues to phrases found in natural languages (Morgan et al., 1987). These features, when present in combination, might make the language easier to learn for a well-equipped learner.

All-combined control. We created a matched control for the all-combined condition. This control language has word classes A through F and uses the assignment of words shown in Table 3. To flatten the transitional probabilities as much as possible, we allowed both phrasal word pairs and nonphrasal word pairs to be optional, repeated, and/or moved. In many ways, we constructed the sentences for this control condition directly from the all-combined condition. The 37 canonical sentences were identical in type and token in both conditions. For the other 37 sentences, we varied whether it was a phrasal or a nonphrasal word pair that was optional, repeated, and/or moved orthogonally across the sentences. In other words, it was not the case that one sentence would have all the features operating over phrasal word pairs and another sentence would have all the features operating over nonphrasal word pairs. In this way, we took each of the 37 noncanonical sentences from the all-combined presentation set and created a similar but slightly altered sentence for the all-combined control presentation set. Example sentences are: ZOR FAL NEB TID (a sentence with the structure DEBC) and TID LUM KOF NAV LEV JES ZOR TAF (a sentence with the structure CDAFBCDE). All 74 sentences were then randomized and concatenated with a 1.6-s isi to form the presentation set, which was looped and played to participants for a total exposure time of approximately 20 min, on each of 5 consecutive days. The transitional probabilities are shown in Table 2.

Procedure

The procedure for Experiment 3 was identical to the procedure for Experiments 1 and 2.

Results

Sentence Test

Our first question was whether participants in the all-combined condition learned the basic word order of the canonical sentence type better than controls. To answer this question, we analyzed the results of the Sentence Test, which are shown in Figure 5. A one-tailed t test on the results from Day 1 showed no difference between the conditions, $t(1, 22) = .57$, $p = .29$, *ns*. The results from Day 5 were similar, $t(1, 22) = .00$, $p = .5$, *ns*. Although learning was strong in both conditions (81% correct on Day 5), no difference emerged between them, indicating that participants in the all-combined condition did not learn the linear structure of the canonical sentence type better than controls.

Phrase Test

In spite of the nonresult on the Sentence Test, we were still interested in whether participants in the two groups induced different types of language structure. Although the all-combined language had clear transitional probability information cueing the existence of phrases, the overall language was complex. But given the strong performance of participants in the experimental groups in Experiment 2 on the Phrase Test, we hypothesized that the learning mechanism that participants were using would be sufficiently robust to allow them to capitalize on (rather than be overcome by) the structured complexity of the language and use it to organize the linguistic input into a hierarchal phrase structure gram-

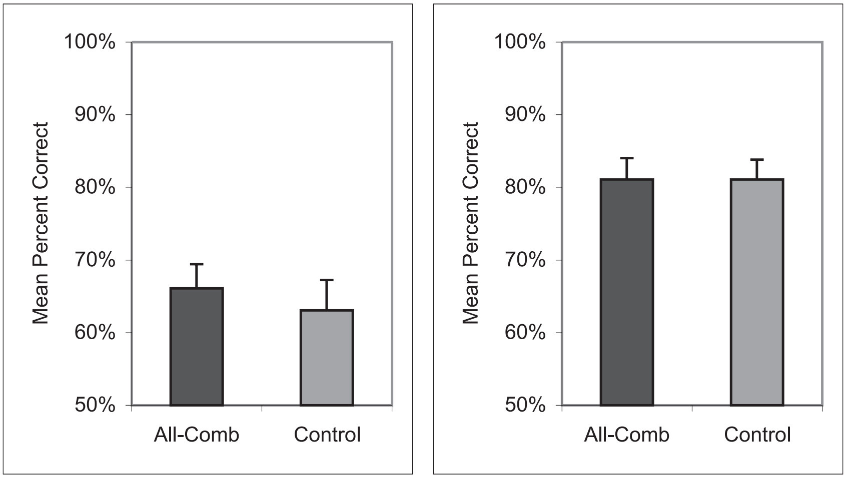


FIGURE 5 Experiment 3: Results of the Sentence Test on Day 1 (left) and Day 5 (right).

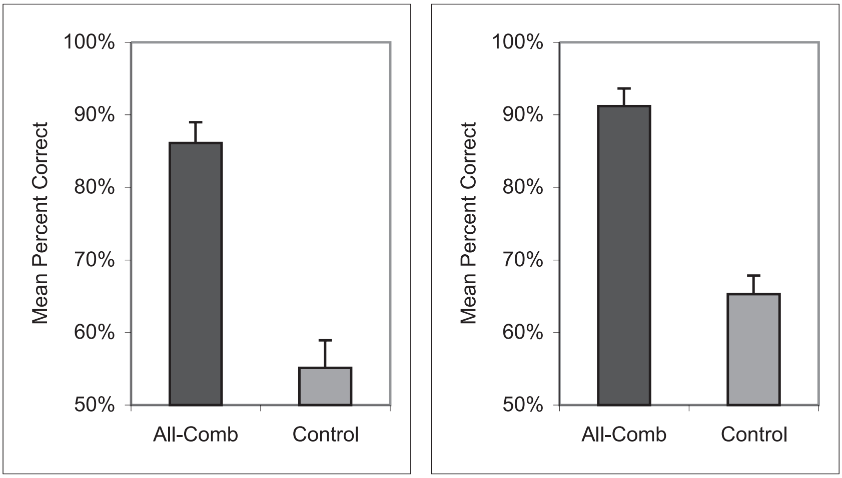


FIGURE 6 Experiment 3: Results of the Phrase Test on Day 1 (left) and Day 5 (right).

mar. To test this hypothesis, we analyzed the results of the Phrase Test, which are shown in Figure 6.

On Day 1, a highly significant difference had already emerged between the two groups, $t(1, 22) = 6.48$, $p = .000$, on a one-tailed t test. This difference remained on Day 5, $t(1, 22) = 7.35$, $p = .000$.

Discussion

The results of the Phrase Test were quite striking. Participants in the all-combined condition learned a hierarchical phrase structure grammar from Day 1, whereas participants in the all-combined control condition did not. On Day 5 the difference between the two groups was still highly significant. However, this difference between the groups did not translate to performance on the Sentence Test. On that test, both groups scored quite well, and no differently from each other. What can we make of this nonresult on the Sentence Test?

It is worth mentioning again that the results of the Sentence Test in Experiment 2 were rather moderate. There was a significant main effect, with experimental groups outperforming controls. However, the difference was small in some cases. In general, the fact that participants in all groups were exposed to so many canonical sentences (50% of all sentences they heard) might have inflated their results on the Sentence Test (which tested only canonical sentences) and masked much of the difference between the groups.

According to this line of reasoning, drastically reducing the number of canonical sentences in the input set should change the pattern of results. Suppose participants were exposed to only a few canonical sentences, say 5% of the total input set (about the same as other sentence types), rather than 50%. Under these conditions, it would be much harder for them to succeed on the Sentence Test by memorizing the linear order of words in the canonical sentence type. As a result, the effect of the experimental manipulation might emerge more strongly on the Sentence Test. Because participants in the experimental conditions, especially those in the all-combined condition, are using transitional probability information to induce a hierarchical phrase structure grammar, in the absence of an abundance of canonical sentences in the input set these participants should be able to succeed on the Sentence Test by using their knowledge of the legal phrasal pairings of words within the language. We tested this hypothesis in Experiment 4.

EXPERIMENT 4

Experiment 4 alters the presentation sets for the all-combined and all-combined control languages so that they contain 5%, rather than 50%, canonical (ABCDEF) sentences. We hypothesized that this change would not affect the strong difference between experimental and control groups on the Phrase Test. However, because it would have been harder for participants to succeed on the Sentence Test by just learning the linear order of words in the canonical sentence type, we hypothesized that participants in the experimental condition would now outperform controls on the Sentence Test, by virtue of their having formed a hierarchical phrase structure representation of the language.

Method

Participants

Twenty monolingual English-speaking undergraduate students were recruited from the University of Rochester to participate in this study. All participants gave informed consent and were paid for their participation.

Description of the Linguistic Systems

The all-combined 5% language. The language for the all-combined 5% condition was identical to the all-combined language from Experiment 3. However, the presentation set had a total of 39 sentences, only 2 of which (or 5%) were of the canonical sentence type. The other 37 sentences combined the features of optional, repeated, and/or moved phrases and were taken directly from the all-combined presentation set from Experiment 3. The 39 sentences were randomized and concatenated with a 1.6-s isi to form the presentation set. The presentation set was looped and played to participants continuously for a total exposure time of approximately 20 min, on each of 5 consecutive days.

Removing 35 of the 37 canonical sentences from the all-combined condition's presentation set had an effect on the transitional probability structure of the input, making the dips in transitional probability at phrase boundaries dramatically more pronounced (see Table 2).

The all-combined 5% control language. The structure of this language was identical to the structure of the all-combined control language from Experiment 3. The presentation set, however, had only 2 canonical sentences (5% of the total number of sentences), plus the additional 37 sentences from the all-combined control condition's presentation set. The 39 sentences were randomized and concatenated with a 1.6-s isi to form the presentation set. The presentation set was looped and played to participants continuously for a total exposure time of approximately 20 min, on each of 5 consecutive days.

Procedure

The procedure for Experiment 4 was identical to the procedure for Experiments 1–3.

Results

Our primary question of interest for this experiment was whether removing the majority (37 out of 39) of the canonical sentences from the presentation sets of the two conditions had the expected impact on performance on the Sentence Test. With much less exposure to ABCDEF sentences, learners could no longer rely on simply learning fixed word order positions to succeed on the Sentence Test. But if learners were capable of succeeding on the Sentence Test by using their knowledge of the phrasal groupings in the language—that is, by learning the word order within phrases and also the relative order of phrases—then participants in the all-combined 5% condition should outperform controls. To test our hypothesis, we analyzed the results of the Sentence Test, shown in Figure 7.

A one-tailed t test on the results from Day 1 showed no difference between experimental and control conditions, $t(1, 18) = .16$, $p = .44$, *ns*. In fact, on Day 1 learners had barely learned the word order in either condition, a result not seen previously. However, by Day 5, a significant difference between the two groups emerged, $t(1, 18) = 2.29$, $p = .035$, with participants in the all-combined 5% condition learning the word order as well as in our previous experiments, presumably by using their knowledge of the hierarchical phrase structure of the language.

To verify that participants in the experimental condition, but not the control condition, had formed a hierarchical phrase structure representation of the lan-

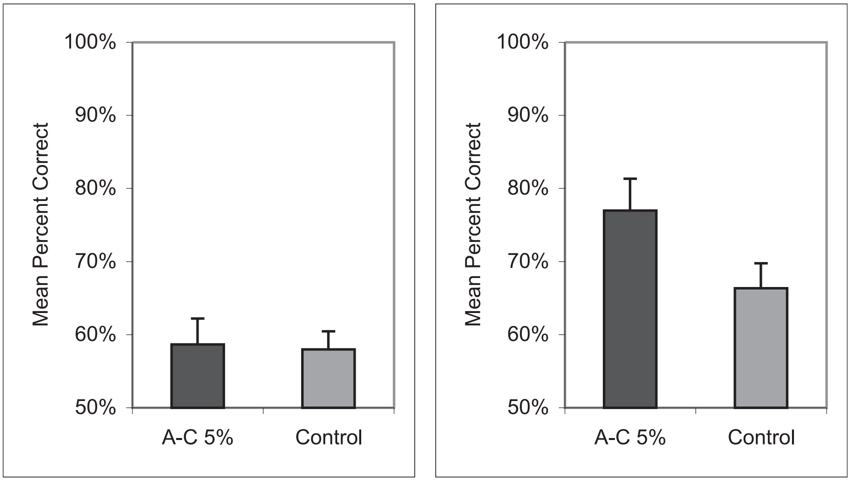


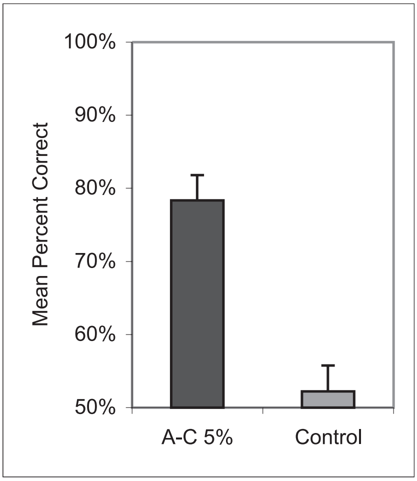
FIGURE 7 Experiment 4: Results of the Sentence Test on Day 1 (left) and Day 5 (right).

guage, we analyzed the results of the Phrase Test, shown in Figure 8. There was a highly significant difference between the two groups on the Phrase Test, both on Day 1, $t(1, 18) = 5.28$, $p = .000$, and on Day 5, $t(1, 18) = 8.41$, $p = .000$, indicating that participants in the all-combined 5% condition had indeed formed a hierarchical phrase structure representation of the language. In contrast (and also in contrast with the control conditions of our earlier experiments), participants in the all-combined 5% control condition did not score above chance on the Phrase Test, suggesting that in the absence of transitional probability cues (and without substantial numbers of ABCDEF sentences), they formed only a flat, finite-state representation of their language.

GENERAL DISCUSSION

Overall, the results of our experiments show that adult learners can use transitional probability peaks within phrases and dips at phrase boundaries to learn the phrases of a miniature artificial language. In addition, our results confirm the findings of previous research (e.g., Morgan et al., 1987, 1989) showing that learning phrases is a necessary step in the comprehensive learning of a miniature artificial grammar.

In Experiment 1 we showed that a simple syntactic feature that is common in natural languages, optional phrases, creates a pattern of peaks and dips in transi-



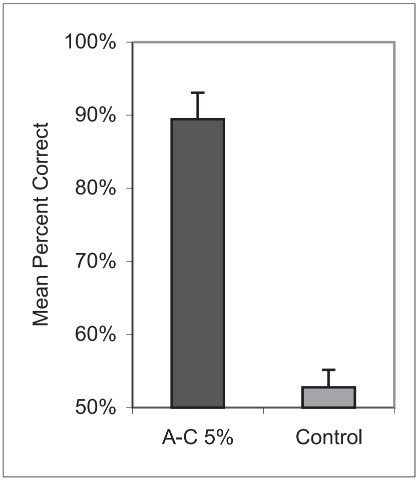


FIGURE 8 Experiment 4: Results of the Phrase Test on Day 1 (left) and Day 5 (right).

tional probability when added to a miniature artificial language. Participants were sensitive to this statistical information and used it to learn the phrases and the overall word order of the language. In Experiment 2 we extended our paradigm to include three more syntactic features that are also common to all natural languages: repeated phrases, moved phrases, and variation in the size of word classes.

In Experiment 3, we asked whether similarly robust learning would be possible if the language were massively more complex than any of the languages in Experiments 1 and 2. We created the all-combined language, which combined in one language all the syntactic features that were tested separately in the first two experiments. The results showed that participants were helped, not hindered, by the increase in complexity. Performance in the experimental conditions on the Phrase Test was striking, even after only 1 day of exposure to the language. However, robust learning on the Phrase Test did not translate into improved performance on the Sentence Test. So in Experiment 4 we drastically reduced the number of canonical sentences in the input with the result that participants in the experimental condition outscored controls, presumably by using their richer, hierarchically structured knowledge of the grammar.

Before discussing the larger significance of these results, we must consider whether there are other explanations, aside from our transitional probability hypothesis for syntax learning, that could account for the pattern of results that we have obtained. Here we consider several other possibilities: the effects of serial position, the use of frequency statistics, and the use of word segmentation strategies.

Serial Position Effects

We are sometimes asked whether the effects of our experimental manipulations could be produced indirectly by serial position effects, rather than by the effects of transitional probability. A number of miniature language experiments, as well as natural language phenomena, have found that learners show better learning of items occurring at the beginnings and ends of sentences than items occurring in the middle (Morgan et al., 1987; Newport, Gleitman, & Gleitman, 1977; Slobin, 1973). Perhaps syntactic operations such as optional or moved phrases increase the frequency or systematicity of word classes appearing in these privileged positions and thereby produce better learning than in control conditions.

There are two responses to this question. First, it is not actually the case that the syntactic operations we have studied do more frequently place word classes in the initial or final positions of sentences. In fact, the control conditions place more different word classes in initial or final position, though some experimental conditions do place certain word classes more consistently in these positions. Second, we have examined the data to see whether there is any tendency for participants in our experiments to show better learning of the initial or final positions of sentences or better identification of the words normally occurring in these positions when they are misplaced in ungrammatical strings. There is no significant difference in any of our experiments between performance on items with errors in the endpoint positions (initial and final) versus middle positions, and also no significant performance difference on items misplacing endpoint items versus middle items. Most important, none of the experiments show a significant interaction between serial position and condition (experimental versus control), which would be required to argue that the effects obtained in these experiments arise from serial position benefits.

Frequency Analysis

We began our experiments with the idea that transitional probability, or an analogous conditionalized statistic, could be used as a cue signaling the existence of phrases in a stream of words. We have suggested that in our experiments adult learners were sensitive to the transitional probability variations in the input, and that this sensitivity was responsible for their learning of phrase structure. Our manipulations do not allow us to distinguish between transitional probability and a number of other related predictive statistics, such as mutual information or conditional entropy, which (like transitional probability) measure the co-occurrence of words or word classes, baselined against the individual frequencies of these elements. One might ask, however, whether the strong performance of experimental groups on the Phrase Test in our experiments was in fact due to the frequency, not the transitional probability, of word sequences. Perhaps participants were sensitive to the frequency with which certain pairs of words appeared together and were not sensitive to transitional probability at all. This is pertinent because the right and wrong answers on the Phrase Test were not counterbalanced for the frequency of those word pairs in the various presentation sets. In other words, in our stimuli, frequency was naturally allowed to covary with transitional probability, as would normally happen in real languages. It is possible to control for frequency and have only transitional probability vary; we have designed such materials within this same paradigm for a future study. But that is a more specific question, which we wanted to ask later.

However, we can ask, in post hoc analyses of the present experiments, whether frequency effects contributed significantly to the present findings. Across items on the Phrase Test there was wide variation in the extent to which the pairs of words that formed the right and wrong answers differed in their frequency of co-occurrence in the presentation sets. On some items, it happened that the wrong answer (the sequence of words that was not a phrase) was nonetheless a sequence of words that appeared more often in a given presentation set than the right answer; on other items, the right and wrong answers appeared equally often; and on still others, the right answer appeared more often. These frequencies differed across conditions, as each condition had a unique set of sentences in the presentation set. For example, SOT FAL versus FAL SIB was Item 1 on the Phrase Test (FAL SIB is the correct answer). In the presentation set for the optional phrases condition, SOT FAL (the wrong answer) appeared four more times than FAL SIB (the right answer). Participants in this condition heard the presentation set 20 times (over 5 days) and therefore heard the wrong answer 80 more times than the right answer; yet all 18 participants chose the right answer on this item on the Phrase Test. In like manner, across items and across conditions, it is possible to analyze whether, in general, participants were discriminating between the two word pairs on a Phrase Test item based on the frequency with which they had heard those words together in their presentation set.

For each condition, we took participants' average scores on each item of the Phrase Test (in the above example, 100%) and paired it with the difference in co-occurrence frequency between the right and wrong answers for that item in that condition's presentation set (in the above example, –4). Then we performed a two-tailed Pearson product moment correlation to see if these two values were correlated. The results are presented in Table 6. Participants' scores showed no significant positive correlation with co-occurrence frequency in any of the conditions from Experiments 1–4, either on Day 1 or on Day 5.[11]

[11]The all-combined 5% control condition does show a significant *negative* correlation on Day 5, indicating that participants chose as the "better group or unit" those pairs of words to which they had been exposed *less frequently* in the input set.

TABLE 6
Pearson Correlations of Frequency Effects on the Phrase Test

Condition	*Day 1*	*Day 5*
Optional phrases	$r(17) = .246$, $p = .324$, *ns*	$r(17) = .245$, $p = .327$, *ns*
Optional control	$r(17) = .429$, $p = .076$, *ns*	$r(17) = .231$, $p = .357$, *ns*
Repeated phrases	$r(17) = .150$, $p = .552$, *ns*	$r(17) = .205$, $p = .414$, *ns*
Repeated control	$r(17) = .345$, $p = .161$, *ns*	$r(17) = .217$, $p = .386$, *ns*
Moved phrases	$r(17) = .119$, $p = .639$, *ns*	$r(17) = .338$, $p = .170$, *ns*
Moved control	$r(17) = .396$, $p = .104$, *ns*	$r(17) = .432$, $p = .073$, *ns*
Class size variation	$r(17) = -.298$, $p = .229$, *ns*	$r(17) = -.070$, $p = .781$, *ns*
Class size control	$r(17) = -.107$, $p = .673$, *ns*	$r(17) = -.056$, $p = .826$, *ns*
All-combined	$r(17) = .024$, $p = .925$, *ns*	$r(17) = .084$, $p = .740$, *ns*
All-combined control	$r(17) = .129$, $p = .609$, *ns*	$r(17) = .176$, $p = .484$, *ns*
All-combined 5%	$r(17) = .057$, $p = .821$, *ns*	$r(17) = .133$, $p = .600$, *ns*
All-combined 5% control	$r(17) = -.182$, $p = .471$, *ns*	$r(17) = -.541$, $p = .021$

One might be surprised that frequency had so little effect on participants' performance. Frequency effects are nearly ubiquitous: they have been demonstrated so often that nearly all psycholinguistic experiments introduce a control for frequency. However, past studies of learning have shown that it is often conditional probability rather than frequency that affects performance. For example, Rescorla (1966) showed that in classical conditioning, it is the conditional probability or predictiveness from a tone to a subsequent shock that affects behavior, whereas the number of times that the tone is followed by the shock has no effect on behavior. In addition, Aslin et al. (1998) demonstrated that 8-month-old infants use conditional probability, in the absence of co-occurrence frequency information, to segment a continuous speech stream into words. Hence the lack of a frequency effect in the present findings is not inconsistent with the learning literature.

Computational Underpinnings

It appears, then, that learners were tracking something like transitional probabilities (or one of its near computational relatives). Our hypothesis is that learners were computing transitional probabilities among word classes[12] and using them to find phrases, as well as to acquire the order of classes within each phrase; the word order of the overall sentence would then be learned as a part of a hierarchical phrase structure representation. If this is correct, it suggests—in combination with

[12]Of course, learners must first (or concurrently) induce the word classes, as these are not transparent in the stream of words to which listeners are exposed. Other researchers have suggested how this might be accomplished via distributional analyses (e.g., Mintz et al., 2002). One possibility is that these processes are interleaved: Learners might initially track the distributions of a small number of individual words, form word classes from these, and then begin to track transitional probabilities among these word classes, while continuing to add individual words to the classes.

our earlier results on transitional probability computations in word segmentation (Aslin et al., 1998; Newport & Aslin, 2000, 2004; Saffran, Newport, & Aslin, 1996)—that such computations are part of a small set of statistical procedures that learners apply at various levels of language acquisition.

However, our results do leave open a wider variety of possibilities. First, learners could have interpreted each sentence as one long, multisyllabic word, and the task at hand as one of word segmentation. Second, learners could have taken each element to be a word (as we intended) but succeeded on the Sentence Test and Phrase Test by implicitly tracking transitional probabilities from word to word, without ever forming word classes.

Word Segmentation Versus Syntax Learning

Throughout this article, we talk about the problem under investigation as one of syntax learning, and we have argued that the same statistical computation used in our earlier studies of word segmentation is here being applied to the problem of syntax acquisition. But is there any possibility that learners in the present experiments are actually treating the task as one of word segmentation?

A number of properties of the exposure materials and the participants' performance suggest that our experiments are tapping syntax learning rather than word segmentation. First, the exposure materials clearly sound like sequences of already-segmented words, not like unsegmented sequences of syllables. Each word is spoken with primary stress, there is minimal coarticulation between words, and there are long pauses between sentences. In addition, for our native English-speaking participants, phonotactic constraints would disfavor combining many of the CVCs into multisyllabic words, such as *hoxjes*. Perhaps most important, as is only the case with syntax, each word belongs to a form class whose members share syntactic privileges; in contrast, if these were syllables forming words, each syllable would have its own specific sequencing privileges. In accord with this word class structure, the exposure sets for each experiment included only a sample of the possible strings, and our participants were tested on novel sequences that they had never heard during exposure.[13] In contrast, our typical word segmentation task presents all the words many times apiece and then tests participants on the specific items that were present in the exposure set. Hence, although both types of learning problems involve inferring the proper way to combine syllable-length elements into longer strings, beneath this superficial similarity there are significant structural differences that make one a problem of word segmentation and the other a problem of syntax learning.

---

[13]Unfortunately, with only three words per class and two classes per phrase, we could not do the analogous procedure in test items for the Phrase Test; to make sure that participants learned the word classes, they were exposed to all legal combinations of words within each phrase. In future studies, it might be helpful to use a larger number of words per class and to leave some of the word combinations for presentation as novel items in the Phrase Test as well.

Computing Statistics Between Words Versus Learning a Phrase Structure

Although our hypothesis concerns the use of transitional probability statistics among word classes to form phrases, our experiments were designed primarily to assess the acquisition of phrases and were not designed to address as thoroughly the contrast between forming phrases from words versus from word classes. However, our data do make some suggestions about the units over which these statistics are being computed.

As we have discussed, the class size variation condition implicitly evaluated whether learners calculate a transitional probability statistic on word classes versus individual lexical items: for this condition alone, a phrasal grouping cue was present in the word-to-word transitional probabilities but not in the class-to-class transitional probabilities. Importantly, the class size variation condition was the only one across our experiments that showed no difference from its control condition on the Phrase Test. There are two possible interpretations of this result. One possibility is that participants in all conditions were computing transitional probability statistics over word classes, and the class size variation condition is the only language in our experiments that did not display its grouping cues at this level of analysis. A second possibility (not previously discussed) is that participants were calculating a word-to-word statistic that combines forward and backward transitional probability (e.g., mutual information). For the class size variation condition, such a statistic is flat across word positions and would therefore not produce a grouping preference stronger than in the control condition.

However, there are two reasons to believe that participants were calculating statistics over form classes. First, as we have mentioned, this is the best explanation for the constellation of facts that the class size variation condition was the only experimental condition that did not outscore its control condition on the Phrase Test, the only experimental condition that did not improve over days on the Phrase Test, and the only experimental condition that would have required an analysis over individual words for participants to succeed on this test. Second, there was a lack of frequency effects on the Phrase Test. At the level of individual lexical items (but not at the level of form classes), bigram frequencies are correlated with bigram transitional probabilities. Calculation of transitional probability over form classes would be the account most consistent with the lack of frequency effects in the data.

The Natural Language Environment

Is the computation of transitional probability likely to be helpful in grouping the words of a natural language into phrases? This is an open question, but we would argue that it would only be helpful if calculated over form classes and not over individual words. First, the data are too sparse if transitional probabilities were calculated over words; many acceptable sequences of words are absent from any reasonable sample. Second, the sizes of form classes are drastically different in natural languages—not different by a factor of 2, as in our experiment, but different by orders of magnitude—perhaps dwarfing any other effects. Even the differences in the frequencies of individual lexical items within a class would affect word-by-word transitional probability statistics too much. Finally, within-phrase and between-phrase transitional probabilities among words can overlap considerably, such that with any individual pair, one would not know which set it belonged to. These issues are alleviated somewhat when the calculation is done over form classes. This dovetails nicely with the evidence (indirect though it is) from the present experiments that participants were forming phrasal groupings by calculating transitional probabilities over form classes.

CONCLUSION

The idea that the formation of phrasal groupings is a critical step in the language acquisition process has a long history in the literature. Morgan, Newport, and colleagues (Morgan & Newport, 1981; Morgan et al., 1987, 1989) showed that external cues to phrase structure, such as prosody, concord morphology, and function words, serve to cue the bracketing of phrasal groupings. In these studies they argued that a rich set of extrasyntactic, correlated cues to phrase structure was necessary to compensate for imperfections in the predictiveness of any one particular cue. Saffran (2001) argued that predictive dependencies within phrases could be an additional cue to phrase structure.

The present findings do not argue against these accounts. Rather, they make the additional suggestion that although extrasyntactic, correlated cues to phrase structure help the learner to bracket phrase groupings from the outside in, intrasyntactic distributional cues can help the learner to bracket phrase groupings from the inside out, via the computation of transitional probability statistics.

The results of the present experiments provide strong evidence that learners are able to calculate transitional probability statistics between adjacent words (or, more likely, word classes) in serially presented sentences, to form phrasal groupings of words based on these statistics, and to use these phrases as an organizing framework within which to better learn the overall structure of the input. This adds to the accumulating evidence that statistical learning may play a role in the acquisition of higher-order levels of language, such as syntax, and suggests a particular type of statistical computation that may apply to syntax as well as to lower levels of language. Our various experiments and conditions each implemented the same underlying statistical pattern, though through quite different syntactic manipulations, and a common pattern of learning was found. Taken together, then, these results support the conclusion that transitional probability exerts a causal influence on performance. Furthermore, these results suggest that a small set of computations may be used to acquire a number of different types of structure or to analyze similar problems at a number of different linguistic levels. One question for further research concerns whether the ability of adults to use transitional probability patterns to form phrasal groupings in the laboratory mirrors processes that subserve the acquisition of syntactic structure in infants. Another question for further research concerns what other importantly different types of computations are needed to handle the full richness of natural language structures. The present results suggest, as hypothesized by Morgan et al. (1987, 1989) and others, that many of the complex properties of syntax in natural languages may function, at least in part, to make such complex languages easier to learn.

ACKNOWLEDGMENTS

This research was supported in part by NIH Grant DC00167 to Elissa Newport, NIH Training Grant DC00035 to the University of Rochester, and NSF Grant SBR-9873477 to Richard Aslin and Elissa Newport.

We thank Dick Aslin, Marie Coppola, Mike Tanenhaus, and Jeff Runner for helpful comments at all phases of this research, and to Susan Goldin-Meadow and three anonymous reviewers for their comments on this article.

REFERENCES

Archer, E. J. (1960). Re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs, 74*(10, Whole No. 497).

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.

Bowerman, M. (1973). Structural relationships in children's utterances: Syntactic or semantic? In T. Moore (Ed.), *Cognitive development and language acquisition* (pp. 197–213). New York: Academic Press.

Braine, M. D. S. (1963). On learning the grammatical order of words. *Psychological Review, 70,* 323–348.

Chomsky, N. A. (1957). *Syntactic structures*. The Hague, The Netherlands: Mouton.

Chomsky, N. A. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Chomsky, N. A. (1981). *Lectures on government and binding.* Dordrecht, The Netherlands: Foris.

Chomsky, N. A. (1995). *The minimalist program.* Cambridge, MA: MIT Press.

Gerken, L. A., Wilson, R., & Lewis, W. (2005). 17-month-olds can use distributional cues to form syntactic categories. *Journal of Child Language, 32,* 249–268.

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development, 1*(1), 23–64.

Gleitman, L. R., & Newport, E. L. (1995). The invention of language by children: Environmental and biological influences on the acquisition of language. In L. R. Gleitman & M. Liberman (Eds.), *An invitation to cognitive science: Vol. 1. Language* (2nd ed., pp. 1–24). Cambridge, MA: MIT Press.

Gleitman, L., & Wanner, E. (1982). Language acquisition: The state of the art. In L. Gleitman & E. Wanner (Eds.), *Language acquisition: The state of the art* (pp. 3–48) New York: Cambridge University Press.

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

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13,* 431–436.

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

Hunt, R. H. (2002). *The induction of categories from distributionally defined contexts: Evidence from a serial reaction time task.* Unpublished doctoral dissertation, University of Rochester, NY.

Jackendoff, R. (1977). *X-bar syntax.* Cambridge, MA: MIT Press.

Lidz, J., Gleitman, H., & Gleitman, L. R. (2003). Understanding how input matters: Verb learning and the footprint of universal grammar. *Cognition, 87,* 151–178.

Marcus, G. F. (2001). *The algebraic mind: Integrating connectionism and cognitive science.* Cambridge, MA: MIT Press.

Maye, J., & Gerken, L. (2001). Learning phonemes: How far can the input take us? In A. H.-J. Do, L. Dominguez, & A. Johansen (Eds.) *Proceedings of the 25th annual Boston University Conference on Language Development* (pp. 480–490). Somerville, MA: Cascadilla Press.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), 101–111.

Miller, G. A., & Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. *American Journal of Psychology*, *63*, 176–185.

Mintz, T. H., Newport, E. L., & Bever, T. G. (1995). Distributional regularities of form class in speech to young children. *Proceedings of North-Eastern Linguistics Society 25* (Vol. 2, pp. 43–54). Amherst, MA: University of Massachusetts, Graduate Linguistic Student Association.

Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science, 26,* 393–425.

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

Morgan, J. L. (1986). *From simple input to complex grammar.* Cambridge, MA: MIT Press.

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,* 67–85.

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

Newport, E. L., & Aslin, R. N. (2000). Innately constrained learning: Blending old and new approaches to language acquisition. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th annual Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48,* 127–162.

Newport, E. L., Gleitman, H., & Gleitman, L. R. (1977). Mother, I'd rather do it myself: Some effects and noneffects of maternal speech style. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition* (pp. 109–149). Cambridge, England: Cambridge University Press.

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

Pinker, S. (1994). *The language instinct: How the mind creates language*. New York: HarperCollins.

Radford, A. (1988). *Transformational grammar, a first course.* Cambridge, England: Cambridge University Press.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 77,* 317–327.

Rescorla, R. A. (1966). Predictability and number of pairings in Pavlovian fear conditioning. *Psychonomic Science, 4,* 383–384.

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

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

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language, 44*, 493–515.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47*, 172–196.

Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science*, 275, 1599–1604.

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. Ferguson & D. Slobin (Eds.), *Studies of child language development* (pp. 489–502). New York: Holt, Rinehart & Winston.

Smith, K. H. (1969). Learning co-occurrence restrictions: Rule learning or rote learning? *Journal of Verbal Learning and Verbal Behavior*, *8*, 319–321.

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

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40*, 374–408.