JOURNAL OF MEMORY AND LANGUAGE **35,** 497–516 (1996)
ARTICLE NO. 0027

Prosody and Adults' Learning of Syntactic Structure

VIRGINIA VALIAN

Hunter College

AND

ANDREA LEVITT

Wellesley College, Haskins Laboratories

The role of prosody in adults' acquisition of a miniature artificial language was examined in three experiments. In Experiment 1, learners heard and repeated prerecorded sentences of the language, and simultaneously saw corresponding referents, but did not see any printed words. Learners received four study-test trials. Half the learners heard a ''single word'' presentation, in which each of the four words of each sentence was recorded with the falling contour associated with list-final position. Half heard a ''phrase prosody'' presentation—expected to aid learning—in which each two-word phrase was recorded as a phrasal unit, with the first two-word phrase of each sentence having a rising contour and the second two-word phrase having a falling contour. Half the participants were given a dialect with high-frequency markers expected to aid learning, and the other half a dialect with low-frequency markers. The phrase prosody presentation did not facilitate learning. Experiment 2 removed the reference field and provided six study-test trials. Phrase prosody here facilitated performance, primarily by increasing learners' acceptance of correct sequences. Experiment 3 removed participants' repetition as well as the reference field and found a strong effect of phrase prosody. We propose that prosody helps recognition of correct word pairs and may be especially useful when other cues to syntactic structure are either unavailable or cannot be exploited by the learner.

The present set of experiments investigates different aspects of the role of prosody in adults' acquisition of a miniature artificial language. Prosody includes cues such as fundamental frequency contour as well as vowel shortening and lengthening, pausing, and loudness. All three of our experiments contrasted phrase prosody with single word prosody using natural speech. In our phrase prosody presentation a rising pitch contour is used on the first two-word phrase of a sentence, and a falling pitch on the second and final two-word phrase. In single word prosody each word is pronounced separately with a falling intonation, as if it were the final item in a list.

While intonation contours are the most noticeable difference between the two types of prosody, temporal patterning and coarticulation effects also distinguish them. It is impossible to maintain naturalness when producing phrases without lengthening the final vowels, thus altering somewhat the relative timing of the two words within a phrase as compared to the single word case. In addition, coarticulation naturally occurs at word boundaries when the words belong to the same phrase or

This research was supported in part by a grant from the National Institute of Child Health and Human Development (HD-24369) to Virginia Valian, in part by a grant from the National Institute of Deafness and Communicative Disorders (DC-00403) to Catherine Best, and in part by a grant from the National Institute of Child Health and Human Development (HD-01994) to Haskins Laboratories. A brief version of this research was presented at the 1991 Psychonomic Society meetings. We thank S. Coulson, A. Reed, and M. Morgon for their assistance in conducting the experiments and D. Mandelbaum and S. Rosenthal for their assistance with data analysis. We thank D. Balota, A. Cutler, R. Meier, M. C. Potter, D. H. Whalen, and three anonymous reviewers for their helpful comments on an earlier draft. Address correspondence and reprint requests to Virginia Valian, Department of Psychology, Hunter College, 695 Park Avenue, New York, NY 10021.

0749-596X/96 $18.00

prosodic unit, whereas no coarticulation occurs for the single word case. Thus, intonation, lengthening, and coarticulation are the dominant cues.

Prosody has been implicated by various investigators as a potentially useful cue for both adult and child listeners. Adult listeners use lengthening and pausing to disambiguate syntactic ambiguities (Lehiste, Olive, & Streeter, 1976; Scott, 1982) and to demarcate phrase boundaries (Harris, Umeda, & Bourne, 1981; Umeda & Quinn, 1981). Such perceptual judgments are affected by fundamental frequency contour as well as other prosodic factors (Beach, 1991; Larkey & Danly, 1982; Lehiste, 1973; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Wales & Toner, 1979; Wales & Taylor, 1987).

Infants can discriminate between utterances on the basis of intonation, timing, or loudness differences (Bull, Eilers, & Oller, 1984, 1985; Eilers, Bull, Oller, & Lewis, 1984; Jusczyk & Thompson, 1978; Spring & Dale, 1977). Infants use prosodic cues to discriminate between utterances of different types, such as declaratives and interrogatives (Best, Levitt, & McRoberts, 1991; Fernald & Kuhl, 1987; Fowler, Smith, & Tassinary, 1985; Jusczyk & Thompson, 1978; Spring & Dale, 1977), and even between their language and another (Mehler, Jusczyk, Lambertz, Halsted, Bertoncini, & Amiel-Tison, 1988).

Prosody also plays a role for older children, who tend to repeat prosodically defined, rather than semantically or syntactically defined, phrases (Reid & Schreiber, 1982) and who understand sentences more easily when they are produced with normal rather than list-like prosody (Tager-Flusberg, 1985). Developmentally, then, prosodic information is one of the first cues used in the perception of language and it continues to be of importance in processing for older children and, to at least a limited degree, adults.

While prosody is attended to by children and adults, its role in syntax learning is unknown. It is plausible to think that infants use prosodic cues as an aid to subsequent syntactic analysis (Kemler Nelson, Hirsh-Pasek, Jusczyk, & Wright Cassidy, 1989; Jusczyk, Hirsh-Pasek, Kemler Nelson, Kennedy, Woodward, & Piwoz, 1992; Morgan, 1986), but there is no direct evidence for such a link, and some have expressed skepticism about a substantial role for prosody in learning about syntax (Crain & Nakayama, 1987; Fernald & McRoberts, 1996; Pinker, 1984). There are limitations to prosody as a structural cue: not all phrase boundaries are prosodically marked, and prosodic phrases are not in 1:1 correspondence with syntactic phrases (Gee & Grosjean, 1983).

Because of the difficulties in studying prosody's role in first language acquisition, the artificial language learning paradigm provides an alternative context in which to investigate whether learners can utilize prosody to work out the structure of a miniature language. Effects of prosody in adults' learning of artificial language cannot imply similar effects for young children acquiring the syntax of their native language. But any obtained effects are a kind of demonstration proof. They show that human learners can utilize prosody in acquiring a structured system.

Two previous studies have found that normal phrase prosody facilitates learning compared either to list-like prosody (Morgan, Meier, & Newport, 1987; Weinert, 1991, cited in Weinert, 1992) or to prosody which was inconsistent with syntactic boundaries (Morgan et al., 1987). Previous work therefore suggests that learners can and will utilize prosody to acquire syntactic structure.

Our experiments expand upon previous work in two important ways. First, our language includes fine-grained dependencies, allowing us to determine whether prosody helps in performing intraphrase analyses, a previously unexamined issue. Second, our experiment embeds prosody within a matrix of other cues, whose values we systematically varied, allowing us to determine whether prosody is an equally strong cue independent of the other cues with which it cooccurs.

Our language was first used by Valian and Coulson (1988). A sentence consists of two distinct two-word phrases, which can appear in either order. The first word of a phrase—which acted as a marker element—is three letters long and begins with a vowel; the second word—which acted as a content word—is four to six letters long and begins with a consonant. The separation of a sentence into two phrases is thus cued by the short–long, short–long pattern of the sentences.

Valian and Coulson (1988) tested the hypothesis that the frequency of the marker relative to the content word would affect whether the learner could acquire the dependency that existed between markers and content words. Only certain pairings were legitimate, but learners' first hypotheses were that any marker could be paired with any content word. In the high-frequency ''dialect,'' each phrase had one marker item and six content words; in the low-frequency dialect each phrase had two marker items and three content words.

Valian and Coulson (1988) found that the high-frequency dialect was much easier to learn than the low-frequency dialect. For both dialects learners quickly established the fact that there were two phrases and that markers preceded content words. But the fine-grained dependency—only certain marker-content pairs were legal—was much easier to establish in the high-frequency dialect. Valian and Coulson suggested that high-frequency markers act as anchor points which facilitate further analysis. When a reference field consisting of one symbol per phrase (rather than one symbol per word) was added, learning was facilitated for both dialects, but the marker–content pairings continued to be easier to learn in the high-frequency dialect.

To summarize, the Valian and Coulson (1988) language had an intraphrase dependency, the learning of which was facilitated both by the presence of high-frequency markers and by the presence of a reference field. The two cues (frequency and reference) provided largely the same kind of information: they identified the two phrases as being of two different types, thereby encouraging learners to search for regularities within each phrase.

As indicated earlier, the second way in which our experiments expand upon previous work is by evaluating the effectiveness of phrase prosody in different matrices of cues. In our experiments, the reference field (when one is included), along with the short–long pattern of the language, cues the learner that there are two different phrases to a sentence. High marker frequency and reference provide cues about marker–content pairings. Despite the fact that high marker frequency and reference supplied very similar information, Valian and Coulson (1988) found that those two cues were additive. Reference did not preempt the effects of marker frequency.

Although previous experiments indicate a facilitating role of prosody, they do not clarify how prosody interacts with other structural cues. In natural spoken language, between-phrase prosodic signals, such as rising and falling intonation and durational lengthening, signal boundary divisions between major phrases. When applied to our artificial language, phrasal intonation would separate a sentence into two phrases, thus overlapping the information given by the short–long, short–long pattern of our language and by the reference field.

In natural spoken language the benefits of prosody within a phrase could include increased internal cohesion via the continuous melodic contour within a phrase and via coarticulation. These features could provide a distinctive additional cue and thus encourage listeners to treat a sequence of words as a unit. They might also benefit learning by promoting a more distinctive and therefore more memorable encoding of phrases, thereby leading to subsequent better recognition of some sequences. But prosody does not uniquely mark phrasal categories as being of a particular type. Noun phrases will receive one prosodic contour in subject position and a different one in object position; one cannot tell that a phrase is a noun phrase via prosody.

Similarly, when applied to our experiments, phrasal intonation gives the phrase ''alt deech'' rising intonation if it is the first phrase of a sentence and falling intonation if it is the last phrase of sentence. Thus, phrasal prosody provides less specific information about the identity of the phrases that are being marked off than does high marker frequency or a reference field.

The present set of experiments examines how different structural cues—frequency, reference, and prosody—interact. We also varied whether learners would repeat each study sentence aloud after listening to it. Since learners of natural language attempt to produce as well as understand language, we tested the possibility that attempting to reproduce the stimulus would draw attention to its structure, its prosodic pattern, or both. If all the cues are additive, then prosody will aid learning regardless of the presence of other cues. But learners might rely on prosody less when other analyzable cues are present, especially if the other cues are more informative or more directly accessible than prosody is. If that is so we would expect prosody to provide the least benefit in learning the high-frequency dialect with a reference field present and the most benefit in learning the low-frequency dialect with no reference field. We thus hypothesize that prosody will be most helpful in the absence of other structural cues.

In all three experiments one group of learners studies the high-frequency dialect and another the low-frequency dialect. A subgroup within each dialect hears a phrase prosody pronunciation and another subgroup hears a list pronunciation. In Experiment 1 learners repeat the study sentences out loud and see a reference field; in Experiment 2 learners repeat the study sentences out loud but have no reference field; in Experiment 3 learners neither repeat the sentences out loud nor see a reference field. If, as we hypothesize, prosody is utilized more as other cues are withdrawn, Experiment 1 should show the weakest effect and Experiment 3 the strongest effect of phrase prosody.

Experiment 1 (Repetition; Reference; Four Trials)

Experiment 1 examines aural learning of a language with visual referents. Learners hear prerecorded sentences of the language but do not see printed words; they do see corresponding referents and repeat the sentences. Experiment 1 is a partial replication of the final experiment of Valian and Coulson (1988): the change is that in this experiment learners hear rather than see the words making up each sentence. Since Valian and Coulson found that the high-frequency dialect of the language was easier to learn than the low-frequency dialect, we expect to replicate that result.

To examine the role of prosody, two different aural presentations were used. In one, the sentences were recorded with a marked phrase prosody, in which the first phrase of each sentence had a rising contour and the second phrase had a falling contour. We refer to this as the phrase prosody condition. In the other, the four words of each sentence were recorded using a single word prosody, with each word having a falling contour. We refer to this as the single word condition.

For our purposes the principal distinguishing features between the two prosodic types are: (a) in the phrase prosody condition the boundary between the two phrases is signalled by the rising contour of the first phrase, while in the single word condition the boundary between words within a phrase is the same as the boundary between the two phrases; and (b) in the phrase prosody condition the words sound slightly different depending on what words they are paired with, because they are coarticulated, while in the single word condition each word sounds exactly the same regardless of what other word it is paired with.

Method

Participants. Forty-eight young adults, Wellesley and MIT students, were paid for their participation. All were native English speakers. Twelve participants were randomly assigned to each of the two prosodic types used in each of the two dialects.

Grammar. The grammar was identical to that used by Valian and Coulson (1988), and the reader is referred to that source for details. Valian and Coulson's description follows.

Each sentence consisted of four words, arranged in two distinct phrases. Schematically, a sentence could take the form [aA bB] or [bB aA], where a lowercase letter represents a marker item and an uppercase letter represents a content item. For example, a sentence could be *alt deech erd hift* or *erd hift alt deech.*

Two dialects of the language were created by manipulating the number of marker and content tokens. The 14-word vocabulary of Dialect 1, the high-frequency dialect, included 2 marker tokens, a single ''a'' and a single ''b,'' and 12 content tokens, 6 A's and 6 B's. The 10-word vocabulary of Dialect 2, the low-frequency dialect, included 4 marker tokens, 2 a's and 2 b's, and 6 content tokens, 3 A's and 3 B's. In absolute terms, the markers of Dialect 1 appeared twice as often as those in Dialect 2. In relative terms, a marker of Dialect 1 appeared six times as often as a given content token, while in Dialect 2 a given marker appeared one and a half times as often as a given content token.

Each dialect had an equal number of strings, 72. In Dialect 1 there were six A-phrases (1 marker × 6 content) and six B-phrases (1 marker × 6 content) and two orders. In Dialect 2 there were also six A-phrases (2 marker × 3 content) and six B-phrases (2 marker × 3 content) and two orders.

All marker items used (a = *alt, ong;* b = *erd, ush*) were one-syllable long and began with a vowel. Thus they supply a phonological correlate, as do many function words in English, to their work as phrase markers. All content items used (A = *deech, tasp, vabe, kicey, logoth, puser;* B = *hift, ghope, skige, cumo, fengle, wadim*) started with a consonant. Words were chosen to avoid sound correspondences or syllable structures that might accidentally make some pairs easier to learn.

Two subdialects were created within each dialect. Dialect 1A used a = *alt* and A = *deech, tasp, vabe, kicey, logoth, puser;* b = *erd* and B = *hift, ghope, skige, cumo, fengle, wadim.* Dialect 1B used a = *ong* and b = *ush,* with the content words arranged as in Dialect 1A. Dialect 2A used a = *alt* and *ong* and A = *puser, tasp, deech;* b = *erd* and *ush* and B = *ghope, hift, wadim;* Dialect 2B used the same markers as in Dialect 2A, and A = *logoth, kicey, vabe,* and B = *cumo, skige, fengle.*

Reference field. The referents used by Valian and Coulson (1988) were used here. For A phrases the referents were colored dot labels 3/4 in. in diameter; the six colors used were yellow, tan, blue, red, orange, and green. For B phrases the referents were six stylized patterns, also round, 1 in. in diameter. The patterns were architectural stamps abstractly representing various shrub and tree patterns. The labels and stamps were horizontally arrayed on otherwise blank 3- × 5-in. index cards. Each card had one colored dot label and one black stamped pattern, about 1 in. apart, with the order determined by the order of the phrases in the training sentences.

Stimuli. For each subdialect 24 study or training sentences and four groups of 24 test strings were prepared. The appendix lists the training sentences for Dialects 1A and 2A. As each training sentence was played the learner was also shown a corresponding index card containing the phrasal referents. No referents were shown with the test stimuli.

Each of the four tests consisted of 12 completely new correct sentences and 12 completely new incorrect strings. Each group of 12 incorrect strings represented four different error types with three examples each. (Valian and Coulson's (1988) description of the stimuli is adapted below.)

Type 1—order errors—violated the requirement that, within a phrase, a marker word precedes a content word. For example, *kicey alt erd cumo* is not a legal string, since *alt* must precede *kicey.* Four forms of incorrect orders [aA Bb], [Aa bB], [bB Aa], and [Bb aA] were used. The incorrect ordering was restricted to a single phrase.

Type 2—double content errors—violated the constraint that a phrase requires both a marker and a content word by replacing the marker with a content word. For example, *erd wadim kicey deech* is not a sentence. *Kicey* and *deech* cannot cooccur within a phrase, even though both are A words. Four forms of incorrect strings, [AA bB], [aA BB], [bB AA], and [BB aA], were used.

Types 3 and *4*—dependency errors—violated the constraint that a marker must appear with a content word of the same phrasal category. In Type 3 the violation appeared only in one of the two phrases. For example, in *alt skige erd wadim, skige* should not cooccur with *alt.* In Dialect 1, all incorrect strings were of the form [aB bB], [bA aA], [aA bA], or [bB aB]. In Dialect 2, the incorrect strings took the forms listed above, of which there were two examples each, and, in addition, took the following forms, of which there was one example each: [aA aB], [bA bB], [bB bA], [aB aA]. *Type 4* strings violated the dependency constraint in both phrases, for example, *alt skige erd deech.* Two forms, [aB bA] or [bA aB], were possible.

Each participant heard four different randomizations of the same 24 training sentences; the same four randomizations were used for participants in each subdialect. The orders of the study sentences were quasirandom, with the constraint that no more than three sequential occurrences of the same marker item could begin a sentence.

For the test items, two orders of each test were recorded with the same constraint and with the additional constraint that at most three responses requiring a ‘‘yes’’ or ‘‘no’’ could appear sequentially. Half the participants received one set of randomizations and half received the other set; randomization was counterbalanced across dialect and prosodic condition. Participants received a different test on each of the four test trials, and across participants four test orders were used, so that each test appeared at least once in each position.

Recording and tape production. All speech was low-pass filtered at 4.9 kHz and digitized at 10 kHz using the Haskins Laboratories PCM system (Whalen, Wiley, Rubin, & Cooper, 1990). For the phrase prosody version, AL produced two versions of each valid marker–content pair. She produced one set of pairs using rising pitch from the marker to the content word, suitable for the first phrase of a two-phrase sequence. She produced the other set using falling pitch from the marker to the content word, suitable for the second phrase of a two-phrase sequence. Within each phrase, due to coarticulation, there was no silent interval between the marker and content word. Via digitization procedures the two phrases of each sentence were created, with a 50-ms silent interval between the two phrases. (To our ears all the 50-ms silent intervals were imperceptible.)

For the single word condition AL pronounced each marker and content word as if it were list final. Valid marker–content pairs were produced by computer with a 50-ms silent interval between the marker and content words within a phrase; the interval was necessary to prevent the marker from sounding abruptly truncated. Via digitization procedures, the two phrases of each sentence were created, with a 50-ms silent interval between the two phrases.

The total average duration of phrase prosody study sentences was 1954 ms for Dialect 1a, 2139 ms for Dialect 1b, 2093 ms for Dialect 2a, and 2123 ms for Dialect 2b. The total average duration of study sentences in the single word presentation was 2329 ms for Dialect 1a, 2345 ms for Dialect 1b, 2340 ms for Dialect 2a, and 2402 ms for Dialect 2b. The single word sentences were thus longer than the phrase prosody sentences. It is unlikely that the shorter duration of the phrase prosody sentences would by itself hinder intelligibility or learning. A 5-s pause was placed between successive study sentences and between the successive test sequences.

For the test sequences the valid sequences were constructed in the same way as the study sentences, using the master versions of each phrase in the phrase prosody condition and the master versions of each marker and content word in the single word condition. The invalid sequences in the phrase prosody condition were constructed by recording each invalid ‘‘phrase’’ with appropriate intonation—either rising or falling intonation. The invalid sequences in the single word condition were constructed by combining the previously recorded items in incorrect combinations.

Procedure. Participants were played an alphabetical list of the words of their subdialect and asked to say each one aloud following the recorded model. Participants were told that they would be presented with 24 sentences from an artificial language that we had made up. They were asked to repeat each recorded training sentence as they heard it. Learners were to try to learn as much as they could about the pattern of the language. Participants were also shown a sample card and told that they would be shown a card for each training sentence they heard and that the pictures on the cards might help in learning the language.

The general outline of the experiment was reviewed. After the study sentences, participants would be played new sequences, one at a time, half of which would be similar to the original sentences and half of which would be different. They were to say yes or no, depending on whether they thought the sentences were like the study sentences. Participants were told they would get no feedback on the correctness of their judgments. They were also told that the new sequences that they would judge would not be accompanied by pictures and that they would not repeat the test sentences out loud. The study-test sequence would be run through four times and would take 45–60 minutes.

Participants were tested individually. The experimenter recorded the learner's responses during the test phases. At the end of the fourth test learners were asked to write down everything they thought they knew about the language. All participants who reached the criterion for learning were also able to give the rules of the language; no participant who failed to reach criterion was able to do so.

Results

As predicted, and consistent with the results of Valian and Coulson (1988), learners of Dialect 1 learned the language better than did learners of Dialect 2. But there were no phrase prosody effects. The participants who heard the phrase prosody versions of Dialects 1 and 2 were no better at learning the language than the participants who heard the single word versions of the dialects.

For all comparisons, errors significantly decreased over trials. Since that result is not of theoretical interest, we do not report the main effect of Trials, only interactions involving Trials.

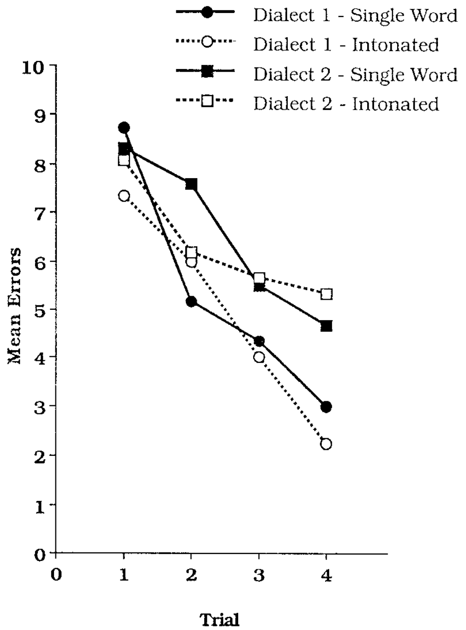

FIG. 1. Experiment 1: Reduction in errors as a function of trial, dialect, and intonation.

Overall data. Figure 1 shows the mean errors as a function of test trial for each prosodic type for each dialect. A three-way analysis of variance (Dialect × Prosody × Trials) revealed a significant effect of Dialect, $F(1,44) = 5.07$, $MSE = 16.3$, $p < .03$. Participants who learned Dialect 1 made fewer errors (M per trial $= 5.10$) than learners who learned Dialect 2 (M per trial $= 6.42$). There was no effect of Prosody ($F(1,44) = .29$), no interaction between Dialect and Prosody ($F(1,44) = .03$), and no other interactions.

A comparison of the number of learners in each of the dialect conditions who mastered the language provides further support for the difference between the two dialects. Learning the language was defined by the criterion of one or fewer errors. Forty-one percent of the Dialect 1 participants learned the language (four from the single word condition and six from the phrase prosody condition) compared to 8% of the Dialect 2 participants.

TABLE 1

EXPERIMENT 1: MEAN ERRORS AS A FUNCTION OF ERROR TYPE, TRIAL, DIALECT, AND INTONATION TYPE

	Trial				
Error type	1	2	3	4	Total
Dialect 1—High frequency, single word					
1—Order	.50	0	0	0	.50
2—Double content	.33	0	.08	.08	.50
3—Single dependency	2.58	2.00	1.50	1.08	7.17
4—Double dependency	2.58	2.00	1.50	1.00	7.08
Total false positive	5.99	4.00	3.08	2.16	15.25
False negative	2.75	1.17	1.25	.83	6.00
Total errors	8.75	5.17	4.33	3.00	21.25
Dialect 1—High frequency, intonated					
1—Order	.33	.08	.08	0	.50
2—Double content	.42	.25	.17	.08	.92
3—Single dependency	2.42	2.08	1.50	1.00	7.00
4—Double dependency	2.17	1.83	1.58	.92	6.50
Total false positive	5.34	4.24	3.33	2.00	14.92
False negative	2.00	1.75	.67	.25	4.67
Total errors	7.33	6.00	4.00	2.25	19.58
Dialect 2–Low frequency, single word					
1—Order	.50	.25	.08	.42	1.25
2—Double content	.58	.58	.17	0	1.33
3—Single dependency	2.25	2.67	1.92	1.92	8.75
4—Double dependency	2.33	2.25	2.67	2.00	9.25
Total false positive	5.66	5.75	4.84	4.34	20.58
False negative	2.67	1.83	.67	.33	5.50
Total errors	8.33	7.58	5.50	4.67	26.08
Dialect 2—Low frequency, intonated					
1—Order	.42	0	0	0	.42
2—Double content	.92	0	.17	.08	1.17
3—Single dependency	2.58	2.58	2.42	2.25	9.83
4—Double dependency	2.17	2.67	2.08	2.33	9.25
Total false positive	6.09	5.25	4.67	4.66	20.67
False negative	2.00	.92	1.00	.67	4.58
Total errors	8.08	6.17	5.67	5.33	25.25

Note. See text (Experiment 1 Method) for full description of Error types. In each trial there are 12 possible false positive and 12 possible false negative. Some totals do not sum because of rounding.

Data by error type. Table 1 presents, for both dialects and both prosodic types, the mean errors for the incorrect strings (learner should say no) and the mean errors for correct strings (learner should say yes). The incorrect strings were false positives and represented four Error Types. Error Types 1 and 2 (order errors and double content errors) were relatively infrequent, and three-way analyses of variance of each (Dialect × Prosody × Trials) showed no main effects of Prosody or Dialect for either Error Type. For Error Type 2 there was a significant interaction of Dialect, Prosody, and Trial, $F(3,132) = 3.59$, $MSE = .19$, $p < .02$: in Dialect 1, error reduction over trials was more rapid in the single word version, while in Dialect 2 it was more rapid in the phrase prosody version.

Error Types 3 and 4 (dependency errors) were more frequent and were analyzed together in a four-way analysis of variance (Dialect × Prosody × Trials × Error Type). In that analysis there was no effect of Prosody, but there was a significant effect of Dialect, $F(1,44) = 7.08$, $MSE = 4.61$, $p < .02$, with Dialect 1 learners making fewer errors. The significant interaction of Dialect and Trials, $F(3,132) = 7.35$, $MSE = .86$, $p = .0001$, indicated a steeper learning curve for Dialect 1.

False negative errors (rejections of correct strings) were also relatively frequent, and were analyzed in a three-way analysis (Dialect × Prosody × Trials). There were no effects of Dialect or Prosody and no interactions.

Discussion

Experiment 1 found the expected benefit of high-frequency markers, but failed to demonstrate any utility of phrase prosody as a cue to structure. There were no effects of phrase prosody either on overall errors or on errors within any error type. The task was also difficult, as indicated by the fact that learners made more errors than comparable learners did in Valian and Coulson (1988), suggesting that a learning task in which the words are purely aural is more difficult for adult learners than one with words presented visually and aurally.

Even under aural presentation, certain aspects of the language appeared easier to learn than others. The test sequences can be divided into two types, those which conform to the short–long, short–long pattern of the language, and those which violate it. The former type includes correct sequences as well as sequences in which incorrect pairings of marker and content words are included (Error Types 3 and 4). In Valian and Coulson (1988), learners were at ceiling from Test 1 on in rejecting sequences that did not conform to the short–long, short–long pattern (Error Types 1 and 2), whether or not reference was present. Here, too, all learners quickly rejected errors that violated the pattern. The short–long, short–long pattern is apparently perceived very quickly, especially in Dialect 1, and additional cues, such as reference and prosody, have little to add.

What was difficult to learn in Valian and Coulson (1988), and even harder to learn here, was that each content word could cooccur only with a particular marker or pair of markers. Learners have difficulty rejecting sequences which conform to the dominant pattern; they reject a certain number of superficially correct sequences simply because they know that some of them must be incorrect. There is thus room in Error Types 3 and 4, and in correct sequences, for prosody to facilitate performance. The puzzle presented by the present results is that phrase prosody neither helped learners reject incorrect pairings nor increased acceptance of correct pairings.

One possible explanation can be rejected. The lack of benefit from prosody cannot be due to its redundancy with reference and high-frequency markers. Recall that Valian and Coulson (1988) found that high-frequency markers and reference were additive cues despite their redundancy.

Prosody may be a cue learners have recourse to only if they have no other cues available or are unable to analyze other cues. Infants' very early sensitivity to prosody occurs at a point when they are unable to analyze syntactic cues. Infants may rely on prosody because they in effect have no other cues to structure. Adults, in contrast, can analyze syntactic cues and may pay less attention to prosodic cues as a result.

The extent of adult listeners' reliance on prosody in normal language comprehension is unclear. Most of the evidence comes from studies of ambiguity resolution. In a recent review paper, Fernald and McRoberts (1996) have sounded a skeptical note concerning the extent to which adults make use of prosody for determining syntactic units. They point out, for example, that listeners make different judgments about what constitutes a sentence when they are supplied with both syntactic and prosodic cues than when they are supplied only with prosodic cues (Geers, 1978; Lehiste, 1979). That is not surprising, since one would expect syntactic cues to signal syntactic boundaries better than prosodic cues. But the findings suggest that in normal comprehension syntax dominates over prosody (Fernald & McRoberts, 1996). The fact that skilled readers can decode prose using only minimal ‘‘prosodic’’ cues, such as commas, may also indicate that prosody is not necessary for language comprehension even if it is helpful. Thus, the adults in our experiments may have learned to disattend to prosody if other structural cues are present.

Recall that Morgan et al. (1987) found facilitatory effects of phrase prosody and did so in the presence of a reference field. One difference between our experiment and theirs is that our reference field clearly cued the existence of two different phrases, while Morgan et al.’s referents cued word boundaries but not phrase boundaries. Our reference field was a grouping cue, while Morgan et al.’s was not. Thus, Morgan et al.’s learners had only prosody to rely on above and beyond the distributional regularities of the language. If prosody is a lower-status structural cue for adults, as we have been speculating, then it would be used if the learner lacked other cues.

To summarize, we found no helpful effects of phrase prosody. Listeners appear not to attend to prosody if they have available a reference field which cues phrase boundaries and phrase identity. Learners’ lack of attention cannot be explained as a general tendency to focus on one cue to the exclusion of others. Learners utilize high marker frequency, but not prosody, in the presence of reference.

EXPERIMENT 2 (REPETITION; NO REFERENCE; SIX TRIALS)

In Experiment 1 phrase prosody was an ineffective structural cue. Learners in that experiment both repeated the sentences they heard and viewed an accompanying reference field. Two alterations were incorporated in Experiment 2.

First, we eliminated the reference field in order to determine whether prosody would now play a more important role. Since we hypothesized that prosody is a lower-status cue compared to reference or high marker frequency, we hypothesized that learners would rely more on prosody if reference were unavailable. Thus, we predicted that prosody would now aid learning, though primarily in Dialect 2, the low-frequency dialect. We also independently predicted that learners would be more successful with Dialect 1, the high-frequency dialect, than Dialect 2, the low-frequency dialect.

The second alteration we made was a consequence of the first. Expecting that removal of the reference field would make the language more difficult to learn, we added two study-test trials, for a total of six.

Method

Participants. Forty-eight young adults, Wellesley and MIT students, were paid for their participation. All were native English speakers. Twelve participants were randomly assigned to each of the two prosodic conditions used in each of the two dialects.

Grammar. The same grammar and strings from Experiment 1 were used in Experiment 2.

Stimuli. The verbal training stimuli were identical to those used in Experiment 1, but the accompanying reference cards were eliminated.

Procedure. Instructions to learners were identical to those from Experiment 1, except that learners were told that the sentences were composed of nonsense words that had no meaning. As in Experiment 1, participants repeated only study stimuli, not test stimuli. In addition, participants were given six study-test sequences. Whichever tests formed tests 1 and 2 for a participant were repeated as tests 5 and 6.

Results

Both major predictions were confirmed. Learners made fewer errors with phrase prosody than single-word intonation and made fewer errors on Dialect 1 than Dialect 2. There was marginal support for our further prediction that phrase prosody would aid Dialect 2

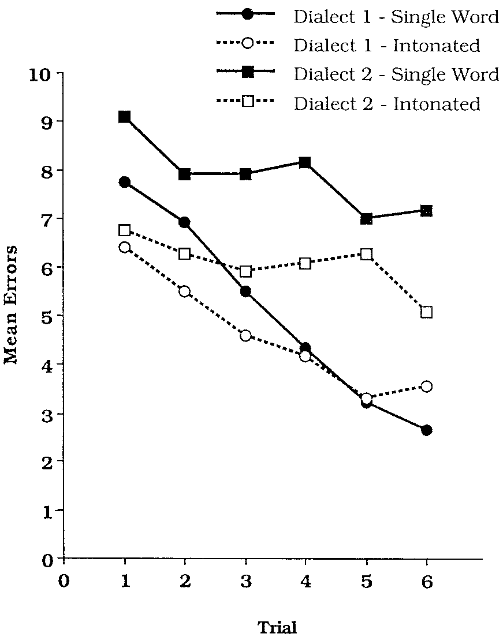

FIG. 2. Experiment 2: Reduction in errors as a function of trial, dialect, and intonation.

more than Dialect 1. As in Experiment 1, Trials was always a significant main effect, and we do not report results for Trials.

Overall data. Figure 2 shows mean errors as a function of test trial (six in this case) for each prosodic type for each dialect. A three-way analysis of variance (Dialect × Prosody × Trials) revealed a significant effect of Prosody, $F(1,44) = 6.54$, $MSE = 14.45$, $p < .02$ learners hearing the phrase prosody version made fewer errors (M per trial $= 5.33$) than learners hearing the single word version (M per trial $= 6.47$).

As predicted, there was also a significant effect of Dialect, $F(1,44) = 22.65$, $MSE = 14.44$, $p < .0001$. Participants who learned Dialect 1 made fewer errors (M per trial $=$ 4.83) than participants who learned Dialect 2 (M per trial $= 6.97$). Twenty-one percent of Dialect 1 learners mastered the language in four trials, and 46% had learned it in six trials. In Experiment 1, 41% of Dialect 1 learners had learned it in four trials, demonstrating the helpfulness of the reference field in Experiment 1. As in Experiment 1, none of the Dialect 2 learners fully mastered the language, even in six trials. There was also a significant interaction between Dialect and Trials, $F(5,220) = 4.55$, $MSE = .90$, $p < .001$. As can be seen in Fig 1, although both groups of participants made fewer errors with each trial, participants learning Dialect 1 tended to show a sharper drop in the number of errors over trials.

There was no interaction across trials between Prosody and Dialect, $F(1,44) = 2.26$, $MSE = 14.45$, $p < .15$. But the predicted interaction did appear in Trial 6, $F(1,44) = 4.33$, $MSE = 6.23$, $p < .05$. Separate analyses showed that there was no difference in Dialect 1 between the phrase prosody errors (M = 3.58) and the single word errors ($M = 2.67$), but there was a significant difference in Dialect 2 between phrase prosody errors (M = 5.08) and single word errors ($M = 7.17$). Phrase prosody aided Dialect 2 more than Dialect 1.

Data by error type. All analyses were performed using six levels of the Trials variable. Table 2 presents, for each dialect and each prosody type, the mean errors for the incorrect strings (learner should say no) and the mean errors for correct strings (learner should say yes). As in Experiment 1, Error Types 1 and 2 (order errors and double content errors) were relatively infrequent. Three-way analyses of variance of each (Dialect × Prosody × Trials) showed a beneficial effect of phrase prosody for Error Type 2, $F(1,44) = 8.06$, $MSE = .50$, $p < .01$, and a beneficial effect of Dialect 1 for Error Type 2, $F(1,44) = 4.02$, $MSE = .50$, $p = .0513$. There were no interactions.

As in Experiment 1, Error Types 3 and 4 (dependency errors) were relatively frequent and were analyzed together in a four-way analysis of variance (Dialect × Prosody × Trials × Error Type). There was no effect of Prosody, but there was a significant effect of Dialect, $F(1,44) = 6.23$, $MSE = 4.50$, $p <$ .02. Dialect 1 learners had fewer errors than Dialect 2 learners. The interaction between Dialect and Trials was significant, $F(5,220) = 5.76$, $MSE = .67$, $p = .0001$. Dialect 1 showed more rapid reduction of errors than Dialect 2.

TABLE 2

Experiment 2: Mean Errors as a Function of Error Type, Trial, Dialect, and Intonation Type

Error type	Trial 1	2	3	4	5	6	Total
Dialect 1—High frequency, single word							
1—Order	.25	.17	.08	.08	0	.08	.67
2—Double content	.42	.17	.33	.25	.17	0	1.33
3—Single dependency	2.33	2.42	2.00	1.75	1.42	1.33	11.25
4—Double dependency	2.58	2.50	1.75	1.25	1.08	1.00	10.17
Total false positive	5.58	5.26	4.16	3.33	2.67	2.41	23.42
False negative	2.17	1.67	1.33	1.00	.58	.25	7.00
Total	7.75	6.92	5.50	4.33	3.25	2.67	30.42
Dialect 1—High frequency, intonated							
1—Order	.17	.08	.17	.17	.25	0	.83
2—Double content	0	.08	.25	.08	.08	.33	.83
3—Single dependency	2.50	2.50	1.83	2.00	1.33	1.42	11.58
4—Double dependency	2.42	2.08	1.67	1.50	1.33	1.33	10.33
Total false positive	5.08	4.74	3.92	3.75	2.99	3.08	23.56
False negative	1.33	.75	.67	.42	.33	.50	4.00
Total	6.41	5.50	4.58	4.17	3.33	3.58	27.57
Dialect 2—Low frequency, single word							
1—Order	.75	.50	.42	.08	.17	.17	2.08
2—Double content	.83	.83	.50	.58	.25	.25	3.25
3—Single dependency	2.25	2.00	2.08	2.25	2.17	2.17	12.92
4—Double dependency	2.08	2.33	2.33	2.25	2.33	2.08	13.42
Total false positive	5.91	5.66	5.33	5.16	4.92	4.67	31.67
False negative	3.17	2.25	2.58	3.00	2.08	2.50	15.58
Total	9.08	7.92	7.92	8.17	7.00	7.17	47.26
Dialect 2—Low frequency, intonated							
1—Order	.33	.08	0	0	.17	0	.58
2—Double content	.42	.25	0	.08	.17	0	.92
3—Single dependency	2.67	2.42	2.42	2.33	2.17	2.33	14.33
4–Double dependency	2.42	2.17	2.25	2.50	2.08	1.83	13.25
Total false positive	5.82	4.92	4.67	4.91	4.59	4.16	29.07
False negative	.92	1.33	1.25	1.17	1.67	.92	7.25
Total	6.75	6.25	5.92	6.08	6.25	5.08	36.33

Note. See text (Experiment 1 Method) for full description of error types. In each trial there are 12 possible false positive and 12 possible false negative. Some totals do not sum because of rounding.

False negative errors (rejections of correct strings) were also relatively frequent and were analyzed in a three-way analysis (Dialect × Prosody × Trials). Prosody emerged as a main effect, $F(1,44) = 8.58$, $MSE = 7.49$, $p < .006$. The phrase prosody version yielded fewer errors than the single word version. Dialect was also significant, $F(1,44) = 9.35$, $MSE = 7.49$, $p < .004$, with Dialect 1 showing fewer errors than Dialect 2.

Comparison between Experiments 1 and 2. Experiment 1 differed from the first four trials in Experiment 2 only in that participants in Experiment 1 saw a reference field as they learned the language. To determine the effectiveness of the reference field we ran a four-way analysis of variance (Experiment × Dialect × Prosody × Trials) on the results of the first four trials in Experiments 1 and 2. There was a marginally significant main effect of Experiment, $F(1,88) = 3.76$, $MSE = 12.24$, $p = .0555$ (Experiment 1 M errors per trial = 5.76, Experiment 2 $M = 6.45$). There was a significant interaction between Experiment and Trials, $F(3,264) = 7.33$, $MSE = 3.69$, $p = .0001$, reflecting the faster learning that took place with a reference field.

There was also a significant main effect of Dialect. Dialect 1 had significantly fewer errors than Dialect 2 (Dialect 1 M per trial = 5.38, Dialect 2 $M = 6.84$), $F(1,88) = 16.81$, $MSE = 12.24$, $p = .0001$. There was a significant interaction between Dialect and Trials, $F(3,264) = 5.32$, $MSE = 3.69$, $p < .01$, reflecting the fact that in general learning is more rapid over trials for Dialect 1.

Finally, despite the complete lack of a prosody effect in Experiment 1, there was a main effect of Prosody for the combined experiments, $F(1,88) = 6.37$, $MSE = 12.24$, $p < .02$, with the phrase prosody version easier than the single word version (phrase prosody M per trial = 5.66, single word M per trial = 6.56).

Discussion

Experiment 2 was a partial replication of Experiment 1, investigating the effects of phrase prosody without a reference field. Our explanation of the lack of effect of prosody in Experiment 1 was that learners relied on reference so that the effects of prosody were masked in Experiment 1. We hypothesized that the effects of phrase prosody emerge most clearly when learners either lack other cues to structure or are unable to utilize an available cue. We thus predicted an effect of prosody in Experiment 2, and prosody did significantly aid learning.

The subanalyses we performed partially clarified the locus of the facilitating effects of prosody: learners who heard phrase prosody accepted correct sequences at a higher rate than learners who heard list prosody. Phrase prosody benefited recognition of correct items and did so equally for both dialects. We propose that prosody helped to make the correct sequences more recognizable by increasing the distinctiveness of each correct pairing. That led to higher acceptance. The greater recognizability that prosody promotes may be a precursor to eventual structural learning (though that did not seem to happen in the present experiment). Noticing that some sequences are repeatedly encountered during training may lead learners eventually to form a rule representing the dependency.

Although the converse of greater recognizability of correct pairings should be lower recognizability and therefore greater rejection of incorrect pairings, that did not occur. We can rule out the possibility that learners hearing the phrase prosody pronunciation were simply adopting a laxer criterion: they accepted correct strings more often in the phrase prosody condition and rejected incorrect pairings at the same rate in phrase prosody and single word prosody.

We suggest instead that correct and incorrect pairings are not genuine converses of each other. Learners tend to accept any string of the short–long, short–long pattern (and then reject some because they know not all can be correct). With correct pairings the greater recognizability which was promoted by phrase prosody added to the tendency to accept correct short–long, short-long sequences, yielding higher acceptance rates. With incorrect pairings the lower recognizability of the incorrect pairs was in conflict with the tendency to accept short–long, short–long sequences. For the participants in Experiment 2, the overall pattern was apparently more dominant in controlling their responses than the lower recognizability of the incorrect pairings.

In this experiment phrase prosody did not facilitate language mastery, which of course requires rejection of incorrect pairings as well as acceptance of correct ones. Almost half the participants of Dialect 1 learned the language by Trial 6, and none of the Dialect 2 participants learned it. Of the 11 Dialect 1 participants who learned the language 6 had received the single word condition and 5 the phrase prosody condition. Thus, in Dialect 1, the presence of high-frequency markers apparently helped learners to reject incorrect pairings, but phrase prosody provided no additional benefit.

We can summarize our conclusions about Experiments 1 and 2 as follows. Phrase prosody benefits learners, primarily by increasing the recognizability of correct strings. There are also limitations to the effects of prosody. The effects are not apparent when a reference field which cues phrase boundaries is present (Experiment 1). Prosody also does not facilitate mastery of dependency learning in a dialect with high-frequency markers (Dialect 1 in Experiment 2). Although reference and high frequency are additive cues, prosody works differently. It does not add to reference at all and appears to add to high frequency in a limited way.

Experiment 3 (No Repetition; No Reference; Four Trials)

Taken together, Experiments 1 and 2 showed that phrase prosody is effective in the absence of a reference field. In Experiment 1 learners were given a reference field and also repeated each study sentence after they heard it: there were no effects of prosody. In Experiment 2 learners did not have a reference field, but did repeat each study sentence. The removal of the reference field resulted in higher acceptance rates for correct strings for learners who heard phrase prosody, regardless of dialect. But phrase prosody did not aid rejection of incorrect pairings; only high-frequency markers improved performance on those error types.

In Experiment 3 we performed a partial replication of Experiment 2, eliminating the cue of repetition. We expected to see the strongest effects of prosody here, because Experiment 3 would give learners the fewest other cues to structure. We used the four trial format of Experiment 1: the relevant prosodic differences appear to be established within four trials. We predicted that phrase prosody would benefit learning, as would the presence of high-frequency markers, in keeping with the results of Experiment 2.

Method

Participants. Thirty-two young adults, Wellesley and MIT students, were paid for their participation. All were native English speakers. Eight participants were randomly assigned to each of the two prosody conditions used in each of the two dialects.

Grammar. The grammar and strings from Experiment 1 and 2 were used in Experiment 3.

Stimuli. The training and test stimuli were identical to those used in Experiment 2 (i.e., they did not contain a reference field).

Procedure. Instructions to participants were similar to those in Experiment 2, but learners were not asked to repeat each sentence out loud after hearing it. We also returned to four study-test trials, in order to make the length of the entire session more manageable.

Results

As predicted, phrase prosody benefited learning. Dialect, contrary to prediction, emerged as a factor only in an interaction with phrase prosody. As in Experiments 1 and 2, Trials was always a significant effect, and we report only interactions with Trials.

Overall data. Figure 3 shows the mean errors as a function of test trial for each dialect for each prosodic type. A three-way analysis of variance (Dialect × Prosody × Trials) revealed a significant effect of Prosody, $F(1,28) = 9.65$, $MSE = 14.31$, $p < .005$. Learners who heard the phrase prosody versions of the two dialects made fewer errors (M per trial $= 4.94$) than learners who heard the single word versions (M per trial $= 7.02$).

Contrary to prediction, there was no main effect for Dialect. Dialect 2 learners in the phrase prosody condition performed anomalously well, suggesting a sampling error. There was a two-way interaction between Prosody and Dialect, $F(1,28) = 4.13$, $MSE = 14.31$, $p = .0517$. Phrase prosody tended to help learners who studied Dialect 2, the low-frequency dialect, more than learners who studied Dialect 1. Two of the 16 participants

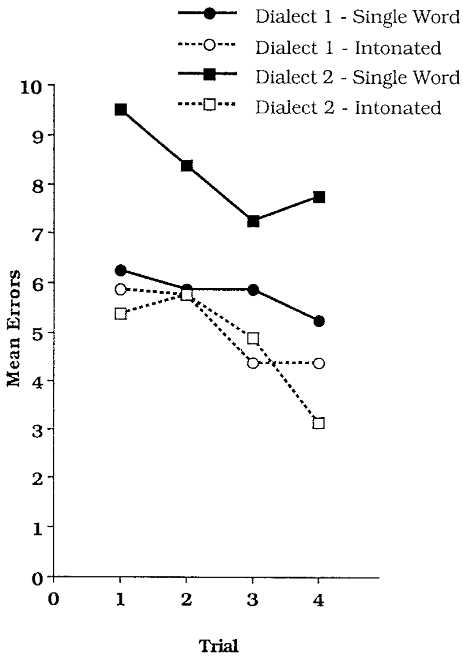

FIG. 3. Experiment 3: Reduction in errors as a function of trial, dialect, and intonation.

in Dialect 1 and 1 of the 16 participants in Dialect 2 learned the language, where learning is defined as ⩽1 error on the last trial.

Data by error type. Table 3 presents, for both dialects and both prosodic types, the mean errors for the incorrect strings (learner should say no) and the mean errors for correct strings (learner should say yes). As in Experiments 1 and 2, Error Types 1 and 2 (order errors and double content errors) were relatively infrequent. Three-way analyses of variance (Dialect × Prosody × Trials) showed no effect of Prosody for Types 1 or 2. Dialect was significant for Type 1, $F(1,28) = 5.16$, $MSE = .80$, $p < .04$, and Type 2, $F(1,28) = 4.17$, $MSE = .48$, $p = .0507$. In both cases, Dialect 1 was easier than Dialect 2.

As in Experiments 1 and 2, Error Types 3 and 4 (dependency errors) were relatively frequent and were analyzed together in a four-way analysis of variance (Dialect × Prosody × Trials × Error Type). Prosody produced a significant main effect, $F(1,28) = 7.33$, $MSE = 3.49$, $p < .02$, with the phrase prosody version showing fewer errors ($M = 1.66$) than the single word version ($M = 2.29$). The interaction between Prosody and Trials was significant, $F(3,84) = 2.85$, $MSE = .83$, $p < .05$. Errors decreased more rapidly over trials for the phrase prosody compared to the single word versions of the dialects. There was no effect of Dialect.

False negative errors (rejections of correct strings) were also relatively frequent and were analyzed in a three-way analysis (Dialect × Prosody × Trials). The interaction between Prosody and Dialect was significant, $F(1,28) = 8.79$, $MSE = 4.48$, $p < .01$. The phrase prosody version of Dialect 2 led to fewer false negative errors than the single word version, but, unexpectedly, the single word version of Dialect 1 led to fewer errors than the phrase prosody version of that dialect. There were no other main effects or interactions.

Comparison between Experiments 2 and 3. The first four trials in Experiment 2 differed from Experiment 3 only in that learners in Experiment 2 repeated the training strings that were presented. To assess the effect of repetition, we ran a four-way analysis of variance (Experiment × Dialect × Prosody × Trials) on the results of first four trials of the two experiments. There was no main effect of Experiment (Experiment 2 M errors per trial = 6.45, Experiment 3 $M = 5.98$), and no interactions involving Experiment.

There were, however, significant main effects for each of the other factors. The phrase prosody version was easier than the single word version (phrase prosody M per trial = 5.32, single word M per trial = 7.11), $F(1,72) = 23.14$, $MSE = 10.56$, $p < .0001$. Further, Dialect 1 had significantly fewer errors than Dialect 2 (Dialect 1 M per trial = 5.55, Dialect 2 M per trial = 6.88), $F(1,72) = 12.88$, $MSE = 10.56$, $p = .0006$. Thus, despite the lack of Dialect effect in Experiment 3, the strong effect in Experiment 2 led to an overall effect when the two experiments were combined. An interaction of Prosody and Dialect, $F(1,72) = 6.50$, $MSE = 10.56$, $p < .02$, showed that the phrase prosody version helped participants

TABLE 3

EXPERIMENT 3: MEAN ERRORS AS A FUNCTION OF ERROR TYPE, TRIAL, DIALECT, AND INTONATION TYPE

	Trial				
Error type	1	2	3	4	Total
Dialect 1—High frequency, single word					
1—Order	0	0	0	.13	.13
2—Double content	.13	0	0	0	.13
3—Single dependency	2.25	2.25	2.25	1.88	8.63
4—Double dependency	2.63	2.38	2.50	2.00	9.50
Total false positive	5.01	4.63	4.75	4.01	18.39
False negative	1.25	1.25	1.13	1.25	4.88
Total errors	6.25	5.88	5.88	5.25	23.27
Dialect 1—High frequency, intonated					
1—Order	0	.13	0	0	.13
2—Double content	.25	0	0	0	.25
3—Single dependency	2.13	1.75	1.50	1.13	6.50
4—Double dependency	1.63	2.00	.88	1.38	5.88
Total false positive	4.01	3.88	2.38	2.51	12.76
False negative	1.88	1.88	2.00	1.88	7.63
Total errors	5.88	5.75	4.38	4.38	20.39
Dialect 2—Low frequency, single word					
1—Order	.88	.75	.50	.50	2.63
2—Double content	.63	.50	.13	.50	1.75
3—Single dependency	2.25	2.50	2.13	2.38	9.25
4—Double dependency	1.63	2.75	2.38	2.50	9.25
Total false positive	5.39	6.50	5.14	5.88	22.88
False negative	4.13	1.88	2.13	1.88	10.00
Total errors	9.50	8.38	7.25	7.75	32.88
Dialect 2—Low frequency, intonated					
1—Order	.50	0	0	0	.50
2—Double content	.13	.25	.13	.13	.63
3—Single dependency	1.63	2.00	2.25	1.13	7.00
4—Double dependency	2.38	2.38	1.38	1.00	7.13
Total false positive	4.64	4.63	3.76	2.26	15.26
False negative	.75	1.13	1.13	.88	3.88
Total errors	5.38	5.75	4.88	3.13	19.14

Note. See text (Experiment 1 Method) for full description of error types. In each trial there are 12 possible false positive and 12 possible false negative. Some totals do not sum because of rounding.

learning Dialect 2 more than it helped participants learning Dialect 1.

Discussion

The results from Experiment 3, where learners simply heard the study sentences, neither repeating them nor viewing a corresponding reference field, show strong beneficial effects of prosody, and the results are very similar to those from Experiment 2. In one difference with Experiment 2, however, phrase prosody here increased only the acceptance rate for correct sequences for the low-frequency dialect, while in Experiment 2 the effect occurred in both dialects. In another difference with Experiment 2, phrase prosody here aided rejection of strings with incorrect marker–content pairings, and did so for both dialects, while in Experiment 2 there was no effect for either dialect. Finally, contrary to prediction, previous findings (Valian & Coulson, 1988), and the findings from Experiments 1 and 2, Dialect 1 was not consistently easier than Dialect 2. We have no explanation for the failure to replicate in Experiment 3, and consider the finding anomalous.

The lack of consistency in the locus of the beneficial effect of prosody should receive further study. In both Experiments 2 and 3 phrase prosody had a strong facilitating effect on learners, but the locus of the effect, both with respect to dialect and with respect to error type, differed. In Experiment 2 prosody led to increased acceptance of correct strings but in Experiment 3 it did not; in Experiment 3 prosody led to increased rejection of incorrect strings but in Experiment 2 it did not.

The extent to which prosody adds to high-frequency markers also requires further study. In Experiment 2 prosody added to the effects of high-frequency markers for recognition of correct sequences but did not add anything to the facilitating effects of high-frequency markers for rejection of incorrect sequences. In Experiment 3 there were no overall effects of high-frequency markers, so that the interaction of prosody and high-frequency markers could not properly be assessed. When Experiments 2 and 3 were analyzed together, phrase prosody and high-frequency markers independently aided learning, but phrase prosody benefitted the high-frequency dialect less than the low-frequency dialect. This suggests that prosodic cues are highlighted in the absence of other structural cues.

General Discussion

Experiments 1–3 investigated the utility of prosody as a cue to learning a miniature artificial language. We contrasted the effectiveness of a phrase prosody pronunciation with a single word pronunciation when prosody was used in combination with three other cues: marker frequency (Experiments 1–3); participant repetition of the stimuli (Experiments 1 and 2); and a reference field which was reliably correlated with the phrasal divisions (Experiment 1). We found that phrase prosody was not utilized by learners at all when our reference field was present and was used to eliminate different kinds of errors depending on the experiment.

We attempted through subanalyses to localize the facilitative effects of phrase prosody. One benefit of phrase prosody was in the recognition of correct sentences. In Experiment 2 both dialects benefited, but in Experiment 3, only the more difficult, low-frequency, dialect was helped. Our explanation of the facilitation is that phrase prosody, including intonation contours along with the coarticulation of the marker and content word within a phrase, made individual phrases more distinctive. Each content word sounded identical in the single word presentation, regardless of what marker it was paired with and regardless of where it appeared in the string.

‘‘Logoth,’’ for example, sounded exactly the same whether correctly preceded by ‘‘alt’’ or incorrectly preceded by ‘‘erd,’’ and whether it was the second or fourth word in the string. In the phrase prosody condition, however, ‘‘logoth’’ sounded slightly different when preceded by ‘‘alt’’ than ‘‘erd,’’ because of the accommodation made in coarticulation, and it sounded different in second position compared to fourth position, because of differences in the pitch contours. That greater distinctiveness, we suggest, produced more memorable encoding of correct pairings and thus led to greater recognizability of those pairings, as reflected in higher acceptance rates for correct sequences. (See related discussion by Speer, Crowder, & Thomas, 1993, who propose that the prosody of a sentence is stored with its syntax and semantics. We are suggesting further that the kind of prosody in which a sentence is presented affects the perceiver’s memory.)

Further facilitation from phrase prosody of within-phrase dependencies occurred in Experiment 3, where phrase prosody learners rejected incorrect pairings at a higher rate than did single word learners. We hypothesize that for learners in both experiments the incorrect pairings sounded less familiar. Also in both experiments, the lesser familiarity of particular pairings competed with the overall great familiarity of the short–long, short–long pattern. In Experiment 2, the overall pattern dominated learners' responses, while in Experiment 3, the unfamiliarity of the incorrect pairings dominated learners' responses.

Since the learners in our experiments are faced with a miniature artificial language instead of a full natural language, and are able to analyze all structural cues, it is difficult to generalize our results to infants. We note again, however, that infants are unable to utilize most of the cues with which they are presented and should therefore rely more on prosody. Our results show clearly that prosody can both increase the recognizability of correct sequences and aid in the rejection of incorrect sequences, even at the relatively fine-grained level of word pairs and even for learners who are sophisticated users of structural cues. Prosody would accordingly be expected to have beneficial effects in first language acquisition (although the ability to reject incorrect sequences is unlikely to be needed).

Our research, considered in the context of previous work, suggests a complex relation between phrase prosody and other structural cues. High marker frequency and reference are additive (Valian & Coulson, 1988). Experiment 1 showed that prosody and reference are not additive. But Experiments 2 and 3 showed that, in the absence of a reference field, prosody is sometimes additive with high-frequency markers. Thus, while our research suggests some limitations to the benefits of prosody, it also demonstrates that the benefits extend beyond the establishment of phrase boundaries. Learners can use prosody to acquire fine-grained aspects of syntactic structure.

Appendix

Study Sentences for Dialects 1a and 2a

Dialect 1a	Dialect 2a
ALT PUSER ERD GHOPE	ERD GHOPE ONG PUSER
ALT PUSER ERD WADIM	USH HIFT ONG PUSER
ALT LOGOTH ERD CUMO	ERD GHOPE ONG TASP
ALT LOGOTH ERD HIFT	USH HIFT ONG TASP
ALT KICEY ERD FENGLE	ERD WADIM ALT DEECH
ALT KICEY ERD HIFT	USH GHOPE ALT DEECH
ALT VABE ERD SKIGE	ERD WADIM ALT TASP
ALT VABE ERD FENGLE	USH GHOPE ALT TASP
ALT TASP ERD GHOPE	USH WADIM ONG DEECH
ALT TASP ERD WADIM	ERD HIFT ONG DEECH
ALT DEECH ERD SKIGE	USH WADIM ALT PUSER
ALT DEECH ERD CUMO	ERD HIFT ALT PUSER
ERD SKIGE ALT PUSER	ONG PUSER ERD WADIM
ERD FENGLE ALT PUSER	ONG PUSER USH GHOPE
ERD SKIGE ALT LOGOTH	ONG TASP ERD WADIM
ERD WADIM ALT LOGOTH	ONG TASP USH GHOPE
ERD GHOPE ALT KICEY	ALT DEECH USH WADIM
ERD CUMO ALT KICEY	ALT DEECH ERD HIFT
ERD CUMO ALT VABE	ALT TASP USH WADIM
ERD HIFT ALT VABE	ALT TASP ERD HIFT
ERD FENGLE ALT TASP	ONG DEECH ERD GHOPE
ERD HIFT ALT TASP	ONG DEECH USH HIFT
ERD GHOPE ALT DEECH	ALT PUSER ERD GHOPE
ERD WADIM ALT DEECH	ALT PUSER USH HIFT

REFERENCES

BEACH, C. (1991). The interpretation of prosodic patterns at points of structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language,* **30,** 644–663.

BEST, C., LEVITT, A., & MCROBERTS, G. (1991). Examination of language-specific influences in infants' discrimination of prosodic categories. *Proceedings of the 12th International Congress of Phonetic Sciences* (Aix-en-Provence, France), **4,** 162–165.

BULL, D., EILERS, R., & OLLERS, D. K. (1984). Infant discrimination of intensity variation in multisyllabic contexts. *Journal of the Acoustical Society of America,* **76,** 1–13.

BULL, D., EILERS, R., & OLLER, D. K. (1985). Infant discrimination of final syllable fundamental frequency in multisyllabic stimuli. *Journal of the Acoustical Society of America,* **77,** 289–295.

CRAIN, S., & NAKAYAMA, M. (1987). Structure dependence in grammar formation. *Language,* **63,** 522–543.

EILERS, R., BULL, D., OLLER, D.K., & LEWIS, D. (1984). The discrimination of vowel duration by infants. *Journal of the Acoustical Society of America,* **75,** 213–218.

FERNALD, A., & KUHL, P. (1987). Acoustic determinants of infant preference for motherese. *Infant Behavior and Development,* **10,** 279–293.

FERNALD, A., & MCROBERTS, G. (1996). Prosodic bootstrapping: A critical analysis of the argument and evidence. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to syntax in early acquisition.* Hillsdale, NJ: Erlbaum.

FOWLER, C., SMITH, M., & TASSINARY, L. (1985). Perception of syllable timing by prebabbling infants. *Journal of the Acoustical Society of America,* **79,** 814–825.

GEE, J. P., & GROSJEAN, F. (1983). Performance and structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology,* **15,** 411–458.

GEERS, A. E. (1978). Intonation contour and syntactic structure as predictors of apparent segmentation. *Journal of Experimental Psychology: Human Perception and Performance,* **4,** 273–283.

HARRIS, M. O., UMEDA, N., & BOURNE, J. (1981). Boundary perception in fluent speech. *Journal of Phonetics,* **9,** 1–18.

HIRSH-PASEK, K., KEMLER NELSON, D. G., JUSCZYK, P. W., WRIGHT CASSIDY, K., DRUSS, B., & KENNEDY, L. (1987). Clauses are perceptual units for young infants. *Cognition,* **26,** 269–286.

JUSCZYK, P. W., HIRSH-PASEK, K., KEMLER NELSON, D. G., KENNEDY, L., WOODWARD, A., & PIWOZ, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology,* **24,** 252–293.

JUSCZYK, P. & THOMPSON, E. (1978). Perception of a phonetic contrast in multisyllabic utterances by two-month-old infants. *Perception and Psychophysics,* **23,** 105–109.

KEMLER NELSON, D. G., HIRSH-PASEK, K., JUSCZYK, P. W., & WRIGHT CASSIDY, K. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language,* **16,** 53–68.

LARKEY, L., & DANLY, M. (1982). Fundamental frequency and sentence comprehension. *MIT-RLE Speech Communication Group Working Papers, 2.*

LEHISTE, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa,* **7,** 107–121.

LEHISTE, I. (1979). Perception of sentence and paragraph boundaries. In B. Lindblom & S. Ohman (Eds.), *Frontiers of speech communication research.* New York, NY: Academic Press.

LEHISTE, I., OLIVE, J., & STREETER, L. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America,* **60,** 1199–1202.

MEHLER, J., JUSCZYK, P., LAMBERTZ, G., HALSTED, N., BERTONCINI, J., & AMIEL-TISON, C. (1988). A precursor of language acquisition in young children. *Cognition,* **29,** 143–178.

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

MORGAN, J. L., MEIER, R. P., & NEWPORT, E. L. (1987). Structural packaging in the input to language learning: Contributions of intonational and morphological marking of phrases to the acquisition of language. *Cognitive Psychology,* **19,** 498–550.

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

PRICE, P. J., OSTENDORF, M., SHATTUCK-HUFNAGEL, S., & FONG, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America,* **90,** 2956–2970.

REID, C., & SCHREIBER, P. (1982). Why short subjects are hard to find. In E. Wanner and L. R. Gleitman (Eds.), *Language acquisition: The state of the art.* Cambridge, MA: Harvard University Press.

SCOTT, D. (1982). Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America,* **71,** 996–1007.

SPEER, S. R., CROWDER, R. G., & THOMAS, L. M. (1993). Prosodic structure and sentence recognition. *Journal of Memory and Language,* **32,** 336–358.

SPRING, D., & DALE, P. (1977). Discrimination of linguistic stress in early infancy. *Journal of Speech and Hearing Research,* **20,** 224–232.

TAGER-FLUSBERG, H. (1985). *How children acquire complex syntax: The interaction of prosodic features and grammatical structure.* Paper presented at the Biennial Meeting of the Society of Research in Child Development, Toronto.

UMEDA, N., & QUINN, A. (1981). Word duration as an acoustic measure of boundary perception. *Journal of Phonetics,* **9,** 19–28.

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

WALES, R., & TONER, H. (1979). Intonation and ambiguity. In W. C. Cooper & E. C. T. Walker (Eds.), *Sentence processing.* Hillsdale, NJ: Erlbaum.

WALES, R., & TAYLOR, S. (1987). Intonation cues to questions and statements: How are they perceived? *Language and Speech,* **30,** 199–211.

WEINERT, S. (1992). Deficits in acquiring language structure: The importance of using prosodic cues. *Applied Cognitive Psychology,* **6,** 545–571.

WHALEN, D., WILEY, E., RUBIN, P., & COOPER, F. (1990). The Haskins Laboratories' pulse code modulation (PCM) system. *Behavioral Research Methods, Instruments and Computers,* **22,** 550–559.

(Received August 23, 1993)
(Revision received July 3, 1995)