JOURNAL OF MEMORY AND LANGUAGE 27, 71–86 (1988)

Anchor Points in Language Learning: The Role of Marker Frequency

VIRGINIA VALIAN AND SEANA COULSON

Wellesley College

We examine the role of markers as anchor points in adult learning of a miniature artificial language, with and without an accompanying reference field. Two dialects of the same language were created, differing only in number of grammatical markers and "content" words. In the high-frequency dialect a given marker occurred six times as often as a given content word, while in the low-frequency dialect a given marker occurred one and a half times as often as a given content word. In Experiment 1, without a reference field, subjects in the high-frequency dialect learned the structure of the language easily, but subjects in the low-frequency dialect learned only superficial properties of the language. In Experiment 2, with a reference field, subjects in both conditions learned, but those in the high-frequency condition learned more quickly. We propose that, with or without a reference field, learners use very high-frequency markers as anchor points for distributional analysis. We discuss the implications of our results for first language learning. © 1988 Academic Press, Inc.

How do children first start learning about the syntactic structure of their language? Valian (1986) has demonstrated that by age 2; 6 children are surprisingly knowledgeable about syntactic categories and skilled at distributional analysis. By that age the child's speech conforms, to a large degree, to the distributional patterns of the language. Very young children know that Determiners, Adjectives, and Nouns occur in that order within a phrase, that Determiners appear in environments distinct from Adjectives, that a multiword Noun Phrase is a constituent which the word *it* can substitute for, and so on. Such cases are examples of how morphemes of different categories are distributed within a sentence. Exactly how the child first begins to perform a distributional analysis to acquire knowledge of lexical and phrasal categories is unknown.

We propose that the distributional analysis children engage in is a form of pattern learning. While language consists of domain-specific entities, such as syntactic categories, we propose that children can acquire some information about them, such as how categories pattern in the target language, by a domain-general process like distributional analysis. In the discussion we will consider the issue of whether children and adults perform distributional analyses similarly enough to justify drawing parallels between the research reported here and the natural language learning situation.

The research described here explores the hypothesis that extremely high-frequency lexical items can serve as anchor points for distributional analysis if those morphemes are reliably associated with a particular constituent or category. In English, morphemes fitting that description include *the*, *a*, and *-ing*. Such morphemes are often re-

This research was performed while the first author was Henry R. Luce Professor and the second author a Dana Foundation Fellow at Wellesley College; we gratefully acknowledge the support of the Henry R. Luce Foundation and the Dana Foundation. For her assistance with every aspect of the study, we thank Alynn Harvey. For their helpful discussion, comments, and suggestions, we thank M. C. Potter, J. J. Katz, T. G. Bever, G. Dell, R. Aslin, R. Meier, J. Morgan, and A. Levitt. We also thank the University of Rochester for its support of the first author during manuscript preparation. Correspondence and reprint requests should be addressed to Virginia Valian, Department of Psychology, Hunter College, 695 Park Avenue, New York, NY 10021.

0749-596X/88 $3.00

ferred to as functors (because they perform grammatical functions), or as markers (because they mark constituent boundaries). The hypothesis predicts that, for English, children will use extremely high-frequency morphemes like *the* as anchor points, and observe what words co-occur with the anchor points. The present study uses a miniature artificial language to examine distributional learning.

In natural language acquisition, attention has been focused less on functors than on content words (primarily Nouns and Verbs), in part because content words are highly stressed in speech (Gleitman & Wanner, 1982), in part because object words have clear referents (Grimshaw, 1981), and in part because very early child speech consists primarily of content words. The first two features might make content words easy to perceive and learn, without, however, contributing strongly to learning about structure. The role played by a reference field in natural and artificial language learning is discussed in the general discussion.

In artificial language learning, the role of markers as cues to constituent structure has been extensively examined. Green (1979) found that a referentially empty phrase structure language was unlearnable if no constituent markers, or inconsistent markers, were used, but learnable if consistent constituent markers were available. His results were replicated, using a different language, by Mori and Moeser (1983). Morgan, Meier, and Newport (1987) examined several different types of cues to constituent structure, in a phrase structure language with a reference field. They examined intonation, concord (via common suffixes on items within the same phrase), and individual markers, and found that learners could make use of each cue, even if, as in concord, English itself did not use the device. Similarly, Meier and Bower (1986) found that subjects learning a language using largely English vocabulary were aided if reliable markers were present. The psychological importance of markers has been demonstrated in children as young as four (Braine, 1966). Braine (1963) proposed that the very high frequency of markers could allow them to serve as foci for associations.

In all previous studies investigating markers, the markers have been of extremely high frequency. In Green's (1979) study, for example, there were four different phrase types, each of which was initiated by its own individual marker. In the effective condition, every time a given phrase type appeared a particular marker initiated it, followed by one of four to six different "content" words. Thus, a given marker occurred much more frequently than a given content word. High frequency alone, however, is insufficient to support phrasal learning, as shown by the fact that when a high-frequency item inconsistently initiated more than one phrase type, learners could not acquire the language.

We propose that distributional learning will be facilitated when there is an anchor point around which to organize lexical material. Two properties will be required for an effective anchor: first, an item must be a true marker, that is, reliably associated with a single structure; second, a marker must be of high frequency relative to the other surrounding items. Note that the two features are independent. In English, for example, the number of Determiners is infinite, since number words are Determiners. Some of those number words are never heard, even though, as Determiners, they are reliably associated with Adjectives and Nouns. They contrast with Determiners like *the* and *a*, which are of extremely high frequency.

The present study investigates whether the relative frequency of a marker affects how easily learners can perform distributional analyses. It uses reliable markers, and varies the frequency of a marker relative to the content words by using one marker in the high-frequency case and two in the low-frequency case. We predict that the presence of high-frequency reliable markers will result in faster learning than the presence of low-frequency reliable markers. In Experiment 1 a referentially empty language is used.

Since some studies, especially by Moeser and her colleagues (Moeser, 1977; Mori & Moeser, 1983), have suggested that learning with a reference field takes place differently from learning without a reference field, in Experiment 2 a reference field is supplied.

EXPERIMENT 1

In Experiment 1 a referentially empty language is used in order to determine whether relative frequency is a relevant variable in learning when there are no other cues to constitutent structure. A modified and simplified version of Green's (1979) grammar was used. A sentence consisted of two distinct phrases, which could appear in either order. Each phrase was two words long, and was initiated by a nonsense marker word and completed by a nonsense "content" word. Schematically, a sentence consisted of [aA bB] or [bB aA], where a lowercase letter represented a marker word and an uppercase letter a content word.

A simple language structure was chosen so that the language could potentially be learned within one experimental session, and so that the string sizes of the two dialects could be equalized while also roughly equalizing vocabulary size. Although the language was simple, it was similar to natural languages in exhibiting "movement" of its two phrases. It thus could not be mastered by adopting a position learning strategy. The "movement" could be used as a cue to the dependency between a marker and content word of a given type.

Marker frequency was manipulated by creating two dialects. In the high-frequency dialect, the 14-word vocabulary consisted of two markers (1 a and 1 b) and 12 content words (6 A and 6 B). In the low-frequency dialect, the 10-word vocabulary consisted of four markers (2 a and 2 b) and 6 content words (3 A and 3 B). A given marker was thereby reliably associated with a content word of the same phrase type in both dialects, and the string size of the language was identical in both dialects. In the high-frequency dialect, however, each marker occurred in every sentence, and occurred six times as often as a given content word. In the low-frequency dialect each marker occurred in half the sentences, and occurred one and a half times as often as a given content word. We predicted that the high-frequency dialect would be easier to learn than the low-frequency dialect.

Method

Subjects. Twenty young adults, primarily Wellesley and MIT students, volunteered or were paid for their participation. All were native English speakers. Ten subjects were randomly assigned to each of the two dialects used in the experiment, five to each subdialect.

All subjects had had some experience with another language; many spoke another language fluently. Post hoc analyses revealed no bias in dialect assignment, and no effect of prior language experience, major course of study, or exposure to linguistics.

Grammar. A sentence consisted of four words, arranged in two distinct phrases. A sentence could take the form [aA bB] or [bB aA], where a lowercase letter represented a marker item and an uppercase letter represented 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, 1 a and 1 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 6 A-phrases (1 marker × 6 content) and 6 B-phrases (1 marker × 6 content) and 2 orders. In Dialect 2 there were also 6 A-phrases (2 marker × 3 content) and 6 B-phrases (2 marker × 3 content) and 2 orders.

All marker items used (a = *alt, ong;* b = *erd, ush*) were 3 letters long; all began with a vowel. All content items used (A = *deech, tasp, vabe, kicey, logoth, puser;* B = *hift, ghope, skige, cumo, fengle, wadim*) were 4–6 letters long and 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 b = *erd,* and all 12 content words; Dialect 1B used a = *ong* and b = *ush,* and all 12 content words. Dialect 2A used all 4 markers and A = *puser, tasp, deech,* and B = *ghope, hift, wadim;* Dialect 2B used all 4 markers and A = *logoth, kicey, vabe,* and B = *cumo, skige, fengle.*

Each sentence was typed on a 3 × 5-in. index card.

Stimuli. For each subdialect 24 training sentences and 4 groups of 24 test strings were prepared. The 24 training sentences were equalized so that (a) there was an equal number of [aA bB] and [bB aA], (b) each content word appeared an equal number of times in second and fourth positions, (c) each marker–content pair occurred an equal number of times, (d) each possible content–content pair across phrases appeared an equal number of times, and (e) no content word occurred more often than any other content word. The Appendix lists the training sentences for Dialects 1A and 2A.

Each of the 4 tests consisted of 12 incorrect strings and 12 correct sentences. The correct sentences were balanced on each test as the training sentences had been. Since the total language size was 72 strings it was possible to have all new correct sentences on each test (24 training + 12 × 4 test = 72).

On each test the 12 incorrect strings represented 4 different error types with 3 examples each. Across all 4 tests and within an error type incorrect strings were balanced as much as possible to meet the constraints described for correct sentences, as well as the constraints described below for each error type. Within a given test, incorrect strings were also balanced as nearly as possible to meet the same criteria. Perfect balancing was impossible, with the result that a given test contained unequal numbers of particular lexical items in particular positions.

Four types of incorrect strings were constructed. Each type contained 12 strings, of which a different 3 appeared on each test. *Type 1* violated the ordering 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 were used: [aA Bb], [Aa bB], [bB Aa], [Bb aA]. Thus, the incorrect ordering was restricted to a single phrase. Strings were balanced so that the misordering appeared equally often in the first and second phrases, equally often in A and B phrases, and equally often with each content word and each marker.

Type 2 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, since *kicey* and *deech* cannot co-occur within a phrase, even though both are A words. Four forms of incorrect strings were used: [AA bB], [aA BB], [bB AA], [BB aA]. The same content word never appeared twice within a phrase, to avoid easy detectability. For the same reason, only content words were used: in Dialect 1 using two marker words in a phrase would have meant using the same marker token twice.

Types 3 and *4* violated the phrasal constraint that a marker had to 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 deech, skige* should not co-occur 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 2 examples each, and, in addition, took the following forms, of which there was 1 example each: [aA, aB], [bA bB], [bB bA], [aB aA]. A given lexical token never appeared twice within a string. Thus, the example [aA aB] might be *alt skige ong tasp,* to test whether subjects realized that *alt* and *ong* were of the same category. The parallel error for Dialect 1 was not used, since it would require repetition of a lexical item.

Type 4 strings violated the phrasal constraint in both phrases. Two forms were possible: [aB bA], [bA aB]. An inadvertent error in constructing correct test items, which was not detected until the end of the experiment, resulted in 13, rather than 12, strings of Type 4 for Dialect 1B. Items were balanced as above.

The order of presentation of training sentences was determined by shuffling the cards for each training block. Similarly, each block of test items was shuffled anew for each subject. Five test orders were used, so that every test appeared at least once in every position, and so that a different test preceded every test (except for the fifth order, which repeated some sequences). In each subdialect each of the five test orders occurred once.

Procedure. Participants were told that they would be presented with sentences from an artificial language that we had made up, composed of nonsense words that had no meaning. They were to try to learn as much as they could about the pattern of the language. They were shown an alphabetical list of the words of their subdialect and told to say each one aloud following the experimenter's model. Subjects were allowed to keep the vocabulary list by them throughout the experiment.

Subjects were told the general outline of the experiment. They were to read each card aloud, taking about 3 s per card, turning the card over after reading it. They would then be shown new cards, one at a time, approximately 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 their judgment. Subjects were told they would get no feedback on the correctness of their judgments. The training-test sequence would be run through four times, and would take about 45 min.

Subjects were tested individually. The experimenter sat opposite the subject at a table. She gave the subject the pack of training cards, shuffling it first, and allowed the subject to control presentation within 2–4 s per card. Then she removed the training stack, brought out the predesignated test stack, and shuffled the stack. She turned the cards up one at a time so that subjects could see the card. She recorded the sentence number placed on the back of the card and the subject's response. The first test stack was then removed, and the original training stack brought out. The cards were shuffled and handed to the subject, who again read each one aloud. The procedure was repeated four times. Subjects were told they could study the vocabulary list while cards were being shuffled.

Subjects were asked to make a yes–no judgment rather than a multiple-choice judgment to minimize how much they could learn from the test itself. In a multiple-choice task hypotheses may be suggested to the subject that she would otherwise not entertain. For example, if faced with choosing between *alt kicey erd wadim* and *erd tasp alt cumo,* both of which superficially fit the pattern of the language, subjects would be encouraged to examine the marker–content pairings to discover a basis on which to distinguish the two strings. We wanted to minimize cluing during the tests and ensure that subjects would do the bulk of their learning during the training blocks; thus, it would correspond more fully to the natural language learning situation. Even with the present procedure, the test might have cued learners that their knowledge of the language was incomplete, since they could tell that they were saying "yes" too often.

At the end of the fourth test subjects were asked to write down everything they thought they knew about the language. Then they were given a stack of cards with one vocabulary item per card and asked to sort the cards into as many piles as they wished, with as many or as few cards per pile as they wished. The first query allowed us to determine what participants explicitly knew about the language; the second allowed us to determine whether subjects might have understood the basic grouping pattern even if they made many errors on the last test. In fact, subjects' responses to the queries were consistent with their yes–no judgments on the final test, so those data are not reported. Finally, participants were asked what other languages they knew and whether they had taken any linguistics courses.

Subjects were then given a written explanation tailored for their subdialect. Subjects' questions were answered.

Results

As predicted, learners of Dialect 1, which had markers of higher relative frequency, learned the language much better than did learners of Dialect 2, which had markers of lower relative frequency. We first present overall data, and then data by error type.

Overall data. Figure 1 shows mean errors as a function of test trial for each dialect. On the first test performance was identical: 6.5/24 errors for Dialect 1, and 6.3/24 for Dialect 2. After Test I the two groups began to diverge. By Test IV Dialect 1 subjects averaged 2.1 errors, and Dialect 2 subjects averaged 5.4 errors. A dialect by trials analysis of variance showed a significant effect of dialect, $F(1,18) = 7.6$, $p = .013$, of trials, $F(3,54) = 18.33$, $p < .001$, and a significant interaction, $F(3,54) = 4.46$, $p < .01$.

In Dialect 1, two subjects were essentially error free (1 or 0 errors) from Test II on, an additional three were error free from Test III on, and an additional two were error free on Test IV. Thus, 70% of Dialect I subjects learned the language perfectly, a conclusion reinforced by the written description after Test IV. For those seven learners, the mean number of errors on the trial before 0–1 errors was reached was 4.1. In Dialect 2, by Test IV, no learner had fewer than 3 errors, and the modal response was 6 errors (50% of subjects).

Data by error type. Table 1 presents mean errors for both dialects separately for correct strings (false negatives) and incor-

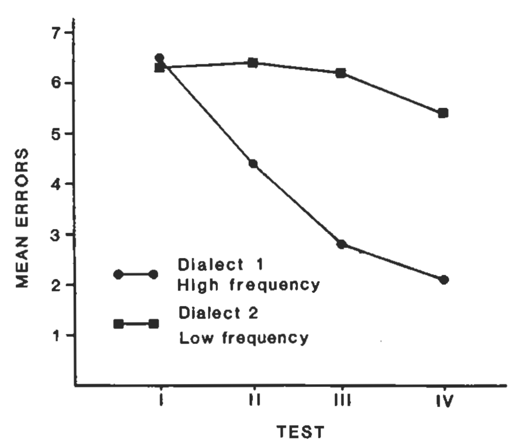

FIG. 1. Experiment 1: Mean errors on each dialect as a function of test trial.

TABLE 1
EXPERIMENT 1: MEAN ERRORS AS A FUNCTION OF ERROR TYPE, TEST BLOCK, AND DIALECT

	Test				
Error type[a]	I	II	III	IV	Total
Dialect 1—High marker frequency					
1	.1	0	0	0	.1
2	.1	0	0	0	.1
3	2.8	2.1	1.3	1.0	7.2
4	2.5	1.4	.7	.6	5.2
All false positive (1–4)	5.5	3.5	2.0	1.6	12.6
False negative	1.0	.9	.8	.5	3.2
Total errors	6.5	4.4	2.8	2.1	15.8
Dialect 2—Low marker frequency					
1	0	0	0	0	0
2	.1	0	0	0	.1
3	2.3	2.8	2.6	2.0	9.7
4	2.4	2.7	2.3	1.7	9.1
All false positive (1–4)	4.8	5.5	4.9	3.7	18.9
False negative	1.5	.9	1.3	1.7	5.4
Total errors	6.3	6.4	6.2	5.4	24.3

Note. For each error type, maximum score = 3; for false negatives, maximum = 12.
[a] For description of error types, see text.

rect strings (false positives), and separately within incorrect strings for each error type. Subjects in both dialects learned immediately that there were two main categories of words, what we have called here marker and content, but which subjects may have represented as short words vs long words, or vowel-initial vs consonant-initial. They also learned immediately that a legal sequence consisted of short–long–short–long. From the first test trial, both groups equally easily rejected Type 1 strings, with order errors (e.g., [Aa bB]), and Type 2 strings, containing two content words within a phrase (e.g., [AA bB]). For both groups rejection was perfect after the first test trial.

The remaining results concentrate on the pattern, over test trials, of false negatives, and false positives for Types 3 and 4. (Types 3 and 4 violated the phrasal constraint by presenting incorrect marker–content pairs.) As Table 1 shows, for false negatives, Dialect 1 subjects showed a slight gradual reduction, from 8 to 4%. For false positives on Types 3 and 4, they showed a marked gradual reduction, from 88 to 27%. In contrast, for false negatives, Dialect 2 subjects first showed a reduction in the percentage of false negatives between Tests I and II, from 13 to 8%, followed by a gradual increase from Tests II to IV, from 8 to 14%. For false positives, they showed first an increase between Tests I and II, from 78 to 92%, and then a gradual reduction from Test blocks II to IV, from 92 to 62%.

An analysis of variance for number of false negatives on the last trial showed a significant effect of dialect, $F(1,18) = 4.53$, $p < .05$. Across all four trials there were no effects of dialect or trial, nor an interaction. An analysis of variance for number of false positives on the last trial showed a significant effect of dialect, $F(1,18) = 6.15$, $p < .03$. Across all four trials there was a significant effect of dialect, $F(1,18) = 8.34$, $p < .01$, of trials, $F(3,54) = 13.94$, $p < .001$, and a dialect by trials interaction, $F(3,54) = 7.25$, $p = .001$.

Discussion

The results strongly support the hypothesis that learners will use high-frequency items as anchor points for distributional analyses of a referentially empty language. Subjects in the low-frequency dialect failed to learn that there were two distinct phrase types, A and B, such that certain content words only appeared with certain markers. Thus, although the language had a simple structure, low-frequency subjects were unable to acquire it in the time available. High-frequency markers appeared necessary and sufficient to learn the marker–content dependency, even without a reference field.

Other structural properties were learned equally well in both dialects. In particular, subjects in both groups immediately learned that legal strings were of the form short–long–short–long (to use the most neutral terminology). Thus, not all relations were impaired, but only those that signaled two types of phrasal constituents.

The pattern of false negatives and false positives suggests that the two groups had two different learning processes. Learners of the high-frequency dialect systematically reduced all errors from the first to last test. Learners of the low-frequency dialect appeared to entrench their short–long hypothesis on the second test, randomly reject a certain number of short–long–short–long sequences on the third, and finally show some improvement on the fourth.

The commonalities and differences between the two groups suggest the language was learned in two stages. In Stage 1 learners in both groups work out the overall short–long–short–long pattern. In Stage 2 learners in the high-frequency group establish the fact of two distinct phrasal categories. Frequency thus appears important in Stage 2 categorical learning.

As mentioned above, previous experiments which have shown strong effects of markers of different types have all contrasted marker presence, in which the marker was of very high frequency, either with marker absence or with a highly frequent item inconsistently initiating phrases of different types. The present results indicate that reliable marker presence is not by itself sufficient to facilitate deeper structural learning. For structural cues to be effective, they must both be reliably associated with a set of items, and highly frequently associated as well. An alternative interpretation of the results, that Dialect 2 presented a more confusing set of associations than Dialect 1, is considered under General Discussion, where a possible mechanism of learning is explored.

It is important to emphasize here that the frequency manipulation was not a syntactic manipulation, but an extrastructural manipulation of linguistic materials. The dialectal difference in learning speed cannot be attributed to structure, since the dialects were structurally identical, and the language sizes were identical. One implication is that a language can be easy or hard to learn, perhaps impossible to learn, for reasons independent of the structure of the target language.

To take this point to its extreme, English (for example) might be unlearnable if some markers (such as *a* and *the*) failed to occur extremely frequently. Our actual linguistic competence, and our acquisition of competence, is mediated by our performance system. That performance system is a composite of representational, acquisitional, analytic, and memorial abilities. As such, it limits us to acquiring a linguistically arbitrary, but cognitively explicable, subset of the natural languages. It even limits us to acquiring a language only under presentation conditions which are cognitively favorable.

The present experiment, in which the language was presented without a reference field, demonstrates the "pure" effects of relative frequency and the extent to which frequency allows markers to function effectively as anchor points. In Experiment 2 we investigate presentation with a reference field, to elucidate the role of frequency in a more realistic language learning context, and to explore the interaction of frequency and reference.

EXPERIMENT 2

In Experiment 2 we added a reference field to determine whether the advantage conferred by high-frequency markers would disappear (see Figure 2 for sample stimuli). We supplied the reference field in a way that would approximate our conception of the natural learning situation. In particular, we supplied subjects with minimal instructions concerning the reference field, and the reference field did not completely duplicate the marker information.

Under certain conditions learners appear to ignore marker information. Mori and Moeser (1983) found that subjects utilized suffix markers when no reference field was supplied, but did not utilize them when a reference field was supplied. They have thus argued that results like Green's (1979), showing that learners can use syntactic

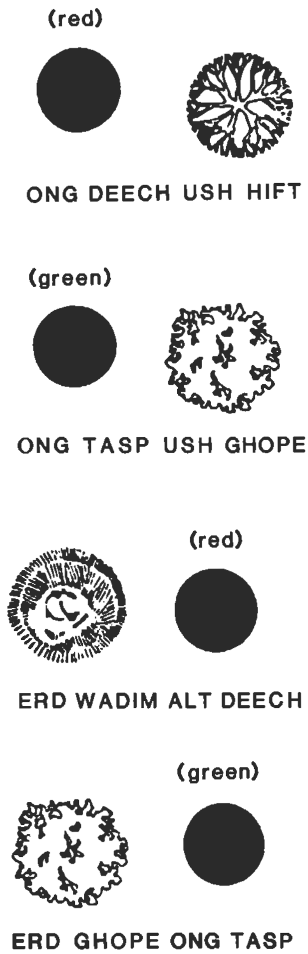

FIG. 2. Experiment 2: Sample stimuli.

markers to acquire complex languages without a reference field, hold only for artificial situations that do not mimic the child's experience. Others (e.g., Morgan et al., 1987), however, have found marker effects even when a reference field is supplied, or have found a language unlearnable without markers even when a reference field is supplied.

Special features of Mori and Moeser's (1983) procedure may have contributed to their findings. In their experiment learners first saw a pictorial display, and then the same display with a sentence underneath it. Subjects were informed that the sentences described the pictures, and were instructed, upon seeing a pictorial display without a sentence, to try to think of the sentence that might describe it. That instruction encourages subjects to search for direct point-for-point correspondences between the pictorial elements and the words in the language. While children may assume a connection between the scenes before them and the words they hear, there is no reason to think they assume a point-for-point connection. With that in mind, we did not include instructions telling subjects that the sentences described the pictures. Pilot work, however, indicated that something had to be said about the pictures, or subjects would be distracted and puzzled by them. Accordingly, we showed subjects a sample training card and informed them that the original (training) cards had pictures on them that might help them learn the language, and that the new strings, where they would be asked to make judgments, would not have any pictures. That minimal information appeared to satisfy subjects.

A second noteworthy feature of Mori and Moeser's (1983) experiment was the "organized reference" condition. Here the syntactic rules of the language governed not only how the words could combine, but also how the pictorial elements could combine. In our view, such a reference field is inappropriate because of its divergence from natural language. In natural language perfect correspondence between syntax and reference, or between syntax and semantics, does not exist. Tense, for example, is not isomorphic with time; it is possible to use the present tense to talk about the present, past, or future time. Similarly, the expressions "the table" and "a table" may both be used to refer to the same table; which of the two expressions is used will depend on nonreferential features like prior linguistic context or assumptions held by the speaker. Finally, some syntactic distinctions, like gender and count/mass, have no invariable referential correlate.

In the critical condition in Mori and

Moeser's (1983) experiment, there was an organized marker system but a reference field that was arbitrarily connected to the different classes of nonsense words. For example, colors, instead of being confined to one class of words, as they were in the organized reference system, were spread across the different classes. In the organized reference system, only one element in a pictured sentence could be colored, while in the arbitrary reference system, more than one could in principle be colored; similarly, only one element could be a rectangle in the organized system, while two rectangles could in principle appear in the arbitrary system. However, the subset of pictorial displays that subjects saw in the arbitrary reference system was chosen so that they were all compatible with the organized reference system. For example, subjects would never be shown a display with two colored elements, or two rectangles, only one. Thus, the displays always appeared to be organized according to some principle. We suggest that subjects tried to determine what that principle was, failed because there was none, and as a result could not learn the language.

The referents in the present experiment included a set of colored dots associated with A phrases and a set of black-stamped stylized patterns associated with B phrases. On each training card a dot was placed above and between the two words of an A phrase and a pattern above and between the two words of a B phrase. Figure 2 displays sample stimuli. Thus, in Dialect 2, *alt deech* and *ong deech* were both associated with the same red dot, and there was no separate visual representation for *alt* or *ong*.

Method

Subjects. Another 20 adults, from the same pool as in Experiment 1, participated in Experiment 2. Ten were assigned randomly to each dialect, 5 to each subdialect.

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

Reference field. For A phrases the referents were colored dot labels ¾ in. in diameter; the six colors 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 colored dot label was placed above and between the two words of an A phrase and the pattern was stamped in black ink above and between the two words of a B phrase.

Stimuli. The training stimuli were identical to those used in Experiment 1, except that the sentence label was placed lower down on the card, with the visual referents above the sentence. The test stimuli were identical to those used in Experiment 1 (i.e., they did not contain a reference field).

Procedure. Instructions to subjects were identical to those of Experiment 1, except that we omitted the statement that the words had no meaning. Instead, subjects were shown a sample training card and told that the original (training) cards had pictures on them that might help in learning the language, and that new strings would not contain pictures.

Results

Overall data. As in Experiment 1, the high-frequency dialect, Dialect 1, was learned more quickly than the low-frequency dialect. Unlike Experiment 1, however, Dialect 2 was learned by some subjects. Figure 3 shows mean errors as a function of test trial. On Test I subjects in both dialects performed equivalently, with 6.7/24 errors for Dialect 1 and 6.3/24 for Dialect 2. As in Experiment 1, subjects learned on the first training trial that there were two main word types, and that they appeared in the order short–long–short–long. On Test II the two groups began to diverge, with Dialect 1 learning faster than Dialect 2. By Test IV Dialect 1 subjects averaged 1.0 errors, and Dialect 2 subjects averaged 3.1 errors. A dialect by trials analysis of variance showed a significant effect of dialect, $F(1,18) = 7.15, p = .016$,

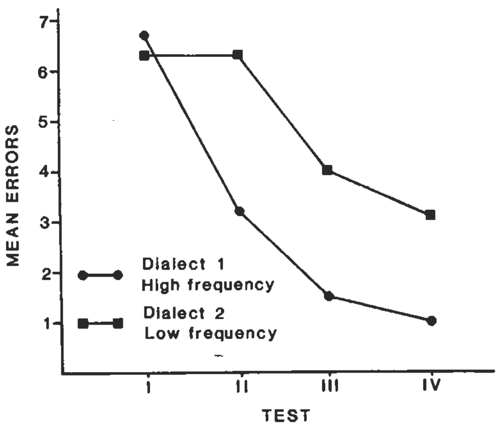

FIG. 3. Experiment 2: Mean errors on each dialect as a function of test trial.

of trials, $F(3,54) = 21.74$, $p < .001$, and a significant interaction, $F(3,54) = 3.17$, $p = .03$. The interaction indicates that Dialect 1 learners began eliminating errors earlier than Dialect 2 learners.

Subjects in Dialect 1 were two trials ahead of subjects in Dialect 2. In Dialect 1, four subjects were essentially error free (1 or 0 errors) as of Test II, eight as of Test III, and seven as of Test IV. (Recall that in Experiment 1 there were seven learners as of Test IV.) The mean number of errors on the trial before 0–1 errors was reached was 5.5. In Dialect 2, two subjects were essentially error free by Test III, and four by Test IV. The mean number of errors on the trial before 0–1 errors was reached was 4.8.

Data by error type. Table 2 presents mean errors for both dialects separately for correct strings (false negatives) and incorrect strings (false positives), and separately within incorrect strings for each error type. As in Experiment 1, learners of both dialects quickly rejected strings violating the short–long–short–long pattern (error types 1 and 2).

Dialect 1 learners gradually reduced false negatives from 11 to 3%, and reduced false positives on phrasal errors from 88 to 10%. Dialect 2 learners averaged 12% false-negative errors, showing slight improvement on Test IV. They reduced false positives on phrasal errors from 72 to 38%, but reduction did not begin until Test III.

TABLE 2
EXPERIMENT 2: MEAN ERRORS AS A FUNCTION OF ERROR TYPE, TEST BLOCK, AND DIALECT

	Test				
Error type[a]	I	II	III	IV	Total
Dialect 1—High marker frequency					
1	0	0	0	0	0
2	.1	0	.1	0	.2
3	2.5	1.2	.3	.3	4.3
4	2.8	1.2	.2	.3	4.5
All false positive (1–4)	5.4	2.4	.6	.6	9.0
False negative	1.3	.8	.9	.4	3.4
Total errors	6.7	3.2	1.5	1.0	12.4
Dialect 2—Low marker frequency					
1	.2	0	0	0	.2
2	.2	.1	0	0	.3
3	2.3	2.5	1.1	1.0	6.9
4	2.0	2.0	1.4	1.3	6.7
All false positive (1–4)	4.7	4.6	2.5	2.3	14.1
False negative	1.6	1.7	1.5	.8	5.6
Total errors	6.3	6.3	4.0	3.1	19.7

Note. For each error type, maximum score = 3; for false negatives, maximum = 12.
[a] For description of error types, see text.

There were no significant effects for false negatives. An analysis of variance for number of false positives on the last trial showed a marginally significant effect of dialect, $F(1,18) = 4.3$, $p < .053$. Across all four trials there was a significant effect of dialect, $F(1,18) = 5.46$, $p < .04$, of trials, $F(3,54) = 23.11$, $p < .001$, and a dialect by trials interaction, $F(3,54) = 4.76$, $p < .006$.

Comparison between Experiments 1 and 2. A dialect by pictures analysis of variance for Test IV errors, comparing Experiments 1 and 2, showed a significant effect of dialect, $F(1,36) = 12.03$, $p < .002$, and of pictures, $F(1,36) = 4.77$, $p < .04$, and no interaction, indicating that the dialect and picture effects were independent of each other. A Newman–Keuls test, with α set at .05, indicated two significant differences on Test IV. Dialect 2 without pictures (M = 5.4) was worse than every other condition, Dialect 2 with pictures (M = 3.1) was worse than Dialect 1 with pictures (M = 1.0). There was no difference between the two Dialect 1 groups (M = 2.1 vs M = 1.0), whether or not pictures were present, and no difference between Dialect 1 without pictures (M = 2.1) and Dialect 2 with pictures (M = 3.1).

Discussion

There are two important features of the results of Experiment 2. First, the addition of a reference field allowed the low-frequency dialect to be learned as well as the high-frequency dialect was learned *without* a reference field. (There was one indication of difference in mastery: seven learners in Dialect 1 without reference obtained 0–1 errors on Test IV, compared to four in Dialect 2 with reference.) A language which would otherwise be unlearnable within four training trials becomes learnable. The presence of a reference field thus compensates for the disadvantage of low-frequency markers.

The second important feature is that, despite the compensatory advantage of a reference field, its presence does not eliminate the frequency effect. The advantage of high-frequency markers is preserved when a reference field is added. Pictures make the low-frequency dialect easier to learn, but they do not overcome the advantage of the high-frequency dialect. The effect of relative high frequency is robust.

It is in Stage 2 categorical learning that a reference field has its facilitative effect. It appears to provide the high-frequency anchor point otherwise lacking in Dialect 2, and to provide an additional very high-frequency anchor for Dialect 1. The data suggest, however, that referents are not exploited as quickly as verbal markers are: high-frequency learners begin eliminating errors on Test II, but low-frequency learners do not begin reducing errors until Test III.

Following Morgan and Newport (1980), we suggest that some reference fields provide effective grouping cues. In the present experiment there was a clear and salient visual distinction between the colored dots on the one hand and the stylized stamps on the other. That constant distinction encouraged learners to form two phrasal categories. In addition, the placement of the visual cue above and between the two words of a phrase may have facilitated phrasal grouping, especially in the low-frequency dialect.

General Discussion

The results of the two experiments strongly supported our prediction that learning an artificial language employing very high-frequency markers would be easier than learning a structurally identical language where the relative frequency of the markers was lower. In the dialect where a given marker was six times as frequent as a given content word, learning was significantly faster and more complete than in the dialect where a given marker was one and one-half times as frequent as a given content word. Those results were obtained whether or not a reference field was present.

What sort of mechanism could explain the results? We have suggested that learners proceed in two stages. In Stage 1, learners try to establish the overall pattern of the language. In this case, the overall pattern consists of the short–long–short–long alternation; it can be acquired rapidly, and independently of marker frequency or reference field presence. In Stage 2, learners search for dependencies in order to establish phrasal categories. In this case, the dependencies consist of marker–content pairings. We suggest that learners track the items which are relatively frequent (the markers), and determine whether there are different privileges of occurrence of the items (the "content" words) which can follow the tracked items. Stage 2 is gradual, and successful only if either high-frequency markers or a reference field is provided.

There are several ways that low-frequency markers could present problems for Stage 2. One possibility is that learners are actively counting frequency of appearance, and do not track items below a certain relative or absolute frequency. Alternatively, high-frequency items may simply be at the top of the lexical look-up stack, and learners try to track the topmost items. In that case, learners who begin tracking all four markers will discover that the same content items appear after more than one marker (e.g., *puser* appears after both *alt* and *ong*). They may prematurely conclude that no disjoint classes exist, and stop tracking. Or learners may find it too difficult to track all four markers, especially since overlap exists among the marker–content pairings, and stop tracking.

In this connection, it should be noted that the pattern of associations within a phrase was different in the two dialects. In the high-frequency dialect, each of the six content words was uniquely associated with a single marker token, while in the low-frequency dialect, each of the three content words was associated with the same two marker tokens. That dual association may have been confusing, suggesting that a good anchor point will have no competitors. If so, the role of frequency may be to rule out competitors.

In any artificial language learning study an important methodological issue is the extent to which findings from artificial language learning can be generalized to natural language learning. On the one hand, the disanalogies are compelling. In addition, in the present case, a very simple language was used, which may not have been analogous to natural language structure, despite our efforts to make it so.

On the other hand, the parallels that exist suggest that some hypotheses concerning natural language acquisition can be usefully explored via artificial language learning. Children do make use of distributional data; distributional analysis is a form of pattern learning; pattern learning is a domain-general process. Thus, it is reasonable to study it artificially as well as naturally, with simple as well as complex languages, and in adults as well as in children. By moving back and forth between the natural and the artificial situations, we can develop and refine hypotheses, and search for relevant evidence in both domains.

For example, our frequency hypothesis predicts both that markers will be of very high frequency in the input to children, and that children will use those markers as anchor points to determine category structure. There are data to support the first prediction. While not an ideal index of input to children, since it uses written text, Kucera and Francis's (1967) analysis shows that half of the 20 most frequent words are Determiners (*the* and *a*, which together account for 9% of all tokens) and Prepositions (*of, to, in, for, with, on, at, by,* which together account for 12% of all tokens). The remainder are other function words; pronouns, some of which function as Determiners; and forms of Modals, *be* and *have*. Thus, the input is likely to contain very high-frequency markers.

The second prediction is more difficult to evaluate, especially since very early child speech lacks many markers. Braine (1963) has suggested that even if children perceive markers simply as noise bursts, such bursts may aid in segmentation of the incoming utterance. There are data to suggest that, whatever the reason for the omission of markers, it is not due to children's failure to perceive them. Children aged 2 and younger use the presence or absence of a Determiner to establish whether a nonsense word refers to a block or a doll/animal (Gelman & Taylor, 1984; Katz, Baker, & Macnamara, 1974). Thus, they perceive the Determiner's sound and assign a function to it. In addition, in an elicited imitation task, 2-year-olds are more accurate in repeating English content words accompanied by English functors than they are in repeating the same content words accompanied by nonsense functors (Gerken, 1987). Thus, children appear sensitive to grammatical markers even if they frequently omit them in speech.

Once children are producing combinatorial speech, the frequency hypothesis would predict that roughly the same markers which are frequent in the input will also be frequent in children's speech. Determiners as markers of Noun Phrases, and Prepositions as markers of Prepositional Phrases, are two obvious marker categories.

In Valian's (1986) analysis of six 2-year-olds' speech, *the* and *a* were the two most frequent Determiners, accounting for 72% of the children's Determiner tokens. Although Valian did not count total tokens, a rough estimate can be obtained by multiplying each child's number of utterances by his or her MLU, and summing. With that estimate, *the* and *a* account for 8% of total tokens. *In, on, of,* and *to* were the children's four most frequent Prepositions, accounting for 71% of their Preposition tokens. In the Kucera and Francis analysis, those four Prepositions were 9% of all tokens, and for the children they were 3% of total estimated tokens. Since Prepositions initiate Prepositional Phrases, and thereby add considerable syntactic complexity, it is impressive that they were as frequent as they were.

The Kucera and Francis data do not, of course, show that such input actually facilitates acquisition, only that it is likely to exist, but our hypothesis would be infirmed if such input did not exist. Similarly, the fact that children's speech makes frequent use of high-frequency markers does not demonstrate that children use such markers as initial anchor points, but their high-frequency usage is predicted by our hypothesis.

The results of the present experiment offer a vantage point from which to examine two major positions on the nature and timing of distributional analysis in the language learning process. There have been striking parallels between theories of natural and artificial language learning. In each, one can distinguish a "referential" position and a "structural" position. Within artificial language learning, the referential position has taken two forms, either that a reference field is necessary if learners are to acquire dependency relations (Moeser, 1977), or that a reference field, when present, is relied on by learners to represent linguistic structure, to the point that learners will ignore linguistic structural cues (Mori & Moeser, 1983). Within natural language learning, the referential position (Grimshaw, 1981; Pinker, 1984; Roeper, 1981) is that the very young child begins language learning with the assumption that all object names are nouns. Whenever an object is named the child then knows that that name is a noun. Such individual nouns serve as the child's entry into syntax. After entry the child notices, for example, that the same word *the* which precedes known object names also precedes nonreferential terms, from which the child deduces that the unknown word is a noun.

All versions of the referential position require the learner to be adept at distributional analysis. The artificial language learning versions have the learner use visual, rather than verbal, categories as the elements on which distributional analysis is performed. The natural language learning versions delay distributional analysis until the child has entered the syntactic system, but do have the child perform the distributional analysis on syntactic categories.

Despite the initial plausibility of referential accounts, closer examination reveals difficulties. Learners of artificial languages acquire formal structure both in the absence of a reference field (Green, 1979; Meier & Bower, 1986; Mori & Moeser, 1983) and in the presence of a reference field (Morgan et al., 1987). As we argued earlier, the only study which has indicated primacy of reference over syntax (Mori & Moeser, 1983) may have achieved its results via the presence of an "unnatural" reference field, namely one which duplicated every feature of the syntax. The present results offer the possibility that a reference field is helpful because it supplies a visual, high-frequency anchor point. If that is correct, a reference field which did not mirror the language's categories would not speed learning, and the presence of high-frequency verbal markers would.

In natural language the reference field does not mirror the language's categories. Linguistic gender, for example, has no invariable referential correlate, but typically there are high-frequency markers signaling gender, and children acquire gender distinctions with relative ease. Further, Valian (1986) argues that if the referential account is intended to apply to children who are producing structured utterances, it should predict confusion among pronouns, nouns, proper names, and noun phrases, since all four categories are used in referring to objects; yet there is evidence that children distinguish those categories by MLU 3. Similarly, Maratsos (1982) points out that children do not confuse adjectives and verbs, despite their referential similarity. Finally, as Morgan et al. (1987) note, work on sign language acquisition (e.g., Newport, 1981; Petitto, in press) indicates that children do not use iconic properties in sign to acquire structure. Thus, referential accounts of language learning seem problematic at best.

Two major problems face structural, nonreferential accounts. The first is how the learner first begins to establish categories. It is to that entry problem that the referential position in natural language learning is addressed. We suggest that the learner establishes categories via extremely high-frequency items. Our hypothesis reinterprets the helpful effects of reference as due to provision of a high-frequency anchor.

The second major problem of distributional accounts is how to constrain the vast number of possible combinations that the distributional analyzer must entertain. Morgan et al. (1987) suggest that many cues to constituent structure serve to localize distributional analysis, in both natural and artificial language learning. If the learner knows where phrase boundaries are, and restricts distributional analyses to units within phrases, or to phrasal units, analysis is feasible. We are in accord with this proposal, adding only that frequency can determine how effective a cue is. Even in a simple language like the one used here, a grouping cue must not only be reliable, but frequent.

Appendix

Training 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

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

BRAINE, M. D. S. (1986). Learning the positions of words relative to a marker element. *Journal of Experimental Psychology*, **72**, 532–540.

GELMAN, S. A., & TAYLOR, M. (1984). How two-year-old children interpret proper and common names for unfamiliar objects. *Child Development*, **55**, 1535–1540.

GERKEN, L. A. (1987). Function morphemes in young children's speech perception and production. Unpublished doctoral dissertation, Columbia University, New York.

GLEITMAN, L. R., & WANNER, E. (1982). Language acquisition: The state of the state of the art. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition* (pp. 3–48). Cambridge: Cambridge Univ. Press.

GREEN, T. R. G. (1979). The necessity of syntax markers: Two experiments with artificial languages. *Journal of Verbal Learning and Verbal Behavior*, **18**, 481–496.

GRIMSHAW, J. (1981). Form, function, and the language acquisition device. In C. L. Baker & J. J. McCarthy (Eds.), *The logical problem of language acquisition*. Cambridge, MA: MIT Press.

KATZ, N., BAKER, E., & MACNAMARA, J. (1974). What's in a name? A study of how children learn common and proper names. *Child Development*, **45**, 469–473.

KUCERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present day American English*. Providence, RI: Brown Univ. Press.

MARATSOS, M. (1982). The child's construction of grammatical categories. In E. Wanner & L. Gleitman (Eds.), *Language acquisition* (pp. 240–266). Cambridge: Cambridge Univ. Press.

MARATSOS, M. P., & CHALKLEY, M. A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K. E. Nelson (Ed.), *Children's language* (Vol. 2, pp. 127–214). New York: Gardner.

MEIER, R. P., & BOWER, G. H. (1986). Semantic reference and phrasal grouping in the acquisition of a miniature phrase structure language. *Journal of Memory and Language*, **25**, 492–505.

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.

MOESER, S. D. (1977). Semantics and miniature artificial languages. In J. Macnamara (Ed.), *Language learning and thought*. New York: Academic 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*, **22**, 498–550.

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

MORI, K., & MOESER, S. D. (1983). The role of syntax markers and semantic referents in learning an artificial language. *Journal of Verbal Learning and Verbal Behavior*, **22**, 701–718.

NEWPORT, E. L. (1981). Constraints on structure: evidence from American Sign Language and language learning. In W. A. Collins (Ed.), *Minnesota Symposia on Child Psychology: Vol. 14. Aspects of the development of competence* (pp. 93–122). Hillsdale, NJ: Erlbaum.

PETITTO, L. (in press). "Language" in the pre-linguistic child. In F. Kessel (Ed.), *The development of language and language researchers*. Hillsdale, NJ: Erlbaum.

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

ROEPER, T. (1981). On the deductive model and the acquisition of productive morphology. In C. L. Baker & J. J. McCarthy (Eds.), *The logical problem of language acquisition*. Cambridge, MA: MIT Press.

VALIAN, V. (1986). Syntactic categories in the speech of young children. *Developmental Psychology*, **22**, 562–579.

(Received April 14, 1987)
(Revision received October 7, 1987)