LEARNING THE POSITIONS OF WORDS RELATIVE TO A MARKER ELEMENT

MARTIN D. S. BRAINE

Walter Reed Army Institute of Research, Washington, D. C.

The learning of conditional definitions of position—"first after g," "second after g"—was studied by exposing children to a subset of the sentences of a miniature, semantically empty, "language." This contained 2 phrase types, fA and gPQ, occurring alone or in either order. Ss learned the structure, including the conditional positional definitions. Errors in sentence-completion test problems indicated a tendency to confuse homologous positions in the 2 phrases, particularly early in learning. The relationships learned could not reasonably be represented by a finite state diagram, and 1 kind of state diagram incorporating a push-down store seemed more appropriate than another. It was argued that finite state diagrams do not capture the properties of associations, particularly remote associations, and that an association might better be regarded as a rule relating contingent items to an entry in a short-term store, or, perhaps, as a context-sensitive rewrite rule.

In previous work (Braine, 1963), a theory of the learning of some aspects of grammatical structure was proposed. The heart of the theory was the notion that "what is learned" are the positions of units in a verbal array, and it was further suggested that units can be hierarchically organized. Thus, people might learn the proper locations of words and morphemes in various sorts of phrases, and what kinds of phrases go in what positions in sentences. In the miniature "languages" used in the experiments, position was always defined absolutely (e.g., "first," "second"). The purpose of the present experiment was to explore the learning of conditional definitions of position, i.e., "before x," "first after x," "second after x," etc., where x is some frequently recurring element which serves, in effect, as a reference point. It may be observed that although conditional relationships of the form "after x" and "before x" pose no obvious theoretical problems—similar kinds of relationships are presumably learned in paired-associate experiments—relationships of the form "second after x" involve a contingency which straddles intervening material. Remote associative connections of this sort have been little studied, although a recent experiment (Braine, 1965b) involved a somewhat similar straddling contingency.

The method used is a modified form of the "verbal reconstructive memory" technique discussed elsewhere (Braine, 1965b; cf. also Smith, 1963, 1965). It differs from the method used in Braine (1963) in that Ss are given no training in the construction of correct rather than incorrect "sentences" under conditions of differential information feedback.

METHOD

Description of the System

The "language" contained two phrase types, A phrases and PQ phrases. There...
were four A phrases (aibi = OB ORDEM, aaba = REMIN GICE, aai = KIVIL, ai = NOOT), which were always preceded by a function word marker (f = YARMO), giving rise to the sequences YARMO OB ORDEM, YARMO REMIN GICE, YARMO KIVIL, YARMO NOOT. PQ phrases also began with a function word marker (g = GED), followed by any of three P words (p1 = MERVO, p2 = YAG, p3 = LECK), followed in turn by any of three Q words (q1 = SOM, q2 = EENA, q3 = WIMP), e.g., GED MERVO SOM, GED MERVO EENA, GED YAG EENA, etc. Complete sentences consisted of either of these phrase types alone ("short" sentences), or of sequences of two phrases, one of each type, in either order ("long" sentences). Thus, the sentence forms were fA, gPQ, fAgPQ, and gPQfA. Since there are four A phrases and nine PQ phrases, the system contains 85 possible sentences, 72 long and 13 short. (Only a subset of these were exposed to S.) It may be observed that no A, P, or Q item is uniquely correlated with an absolute sentence position. However, all A, P, and Q items are uniquely correlated with a particular sentence position defined relative to one or the other marker.

Subjects

The Ss were 20 fourth-grade children, 11 boys and 9 girls, aged 9-1 to 10-6. They were from middle-class professional backgrounds, and of above average intelligence.

Procedure

The S was told that he was going to learn part of a new language which might seem strange because he would not know what any of the words meant. The words, written on cards, were shown to him, and read by E and then by S. The experiment consisted of an exposure period, followed by a test (Test I), another exposure period, and a second test (Test II). The entire procedure occupied a single session of 45-55 min. duration with each S.

Exposure.—During the exposure periods S was exposed to 29 of the 85 possible sentences. The exposed sentences used all four A phrases, and five of the nine possible PQ phrases (gPQfA, gPQfA, gPQfA, gPQfA, and gPQfA). There were 9 different short sentences (the 4 fA and the above 5 gPQ sequences), and 20 different long sentences (10 fAgPQ and 10 gPQfA). Thus, of the 40 different long sentences obtainable by recombining the 4 A and 5 PQ phrases in either order, 20 were exposed and 20 were withheld. Both exposure periods used the same 29 sentences and were similarly organized. The sentences were set up in 10 multiples of 7, 5 short and 2 long (1 fAgPQ, 1 gPQfA); in every set of 7 sentences each of the 4 A and 5 PQ phrases occurred once, either as a short sentence, or as part of a long sentence. Altogether, each exposure period presented the same 70 sentences, 20 long (all different) and 50 short (many repetitions). Each of the 9 A or PQ phrases used occurred 10 times, sometimes as a short sentence, sometimes as part of a long sentence.

The S first read 14 sentences typed on a sheet; if he had difficulty reading, he was helped by E, who reread each sentence after S. The E then read 14 sentences, and S listened and attempted to repeat each sentence after E. If S’s repetition were correct he received a counter; if it were incorrect E repeated the sentence again before going on to the next one. (Sixteen counters were worth one chocolate.) The remaining part of the exposure used a match-from-sample procedure: a sentence was presented and read by S, and then E read two, three, or four sentences, which used the same phrase types as the sample sentence, the last sentence read being the same as the sample sentence. The S had to say for each sentence whether it was the same as, or different from, the sample sentence. The S received a counter if his judgments were correct, which they always were. Forty-two sentences, 14 sample sentences and 28 choice sentences different from the sample, were presented. (The match-from-sample procedure was employed only because it provided a convenient way of exposing a child to a large number of sentences, under conditions assuring attentive listening and unlikely to provoke boredom.)

Tests.—The purpose of the tests was to investigate the extent to which Ss had learned the pattern properties of the sentences. Thus, they were designed to find out (a) whether Ss had learned the formation rule for long sentences—that they were composed of two short sentences, one of each type, (b) whether S could distinguish which word sequences followed f and which followed g, and (c) whether S had registered the internal structure of the PQ phrase. Each test comprised 20 multiple-choice sentence-completion problems, as follows: (In specifying problems below, the dash indicates the location of the space to be filled in.)

1. Four problems, —fA, fA—, LgPQ—, —gPQ. The two alternatives offered were
always an $fA$ and a $gPQ$ sequence. The $A$ and $PQ$ phrases used were always phrases to which $S$ had been exposed; however, the long sentence resulting from a correct choice was one which $S$ had not previously heard (and differed in Tests I and II). Thus, correct responses presumably indicate that $S$ knows how to construct new long sentences out of familiar short ones, i.e., that he knows that a phrase marked by $f$ must be paired with a phrase marked by $g$.

2. Eight problems, four using short sentences ($f—, f—, g—, g—$), and four using long sentences ($gPQf—, gPQf—, g—fA, fAg—$). The short sentences were ones used in the exposure, but the long sentences were new combinations of familiar phrases. The two alternatives offered were always an $A$ and a familiar $PQ$ phrase, a different phrase being correct in each problem. (The phrases correct in the short-sentence problems in Test I were correct in the long-sentence problems in Test II, and vice versa.) These problems test $S$'s knowledge of which sequences follow $f$ and which follow $g$, and whether he is able to make this discrimination in both long and short sentences.

3. Four problems, in which the sentences lacked a word from the $PQ$ phrase. In Test I, the problems were $gPQ—, g—qA, fAgPQ—, g—qfA$; in Test II they were $gPQ—, g—qA, gPQf—, fAg—qA$. Three alternatives were always offered to complete the sentences: a $P$ word (either $p_1$ or $p_2$), a $Q$ word ($q_1$ or $q_2$), and a word from one of the $A$ phrases. (Note that a correct choice always led to the construction of a familiar $PQ$ phrase.) In one short-sentence problem and one long-sentence problem, the $A$ alternative was a word that occupied the position in the $A$ phrase homologous to the vacant position in the $PQ$ phrase in the problem presented (i.e., $a_1$ and $a_2$ when a $P$ word was to be filled in, $b_1$ and $b_2$ for a $Q$ word); in the other problems the $A$ alternative was a nonhomologously positioned word. (Note that a tendency to respond on the basis of absolute position should lead to errors involving a choice of a homologously positioned $A$-phrase item.) These and the next set of problems test $S$'s knowledge of the contents of the $PQ$-phrases.

4. Four problems, similar to Type 3, but employing $p_2$ and $q_2$: $gPQ—, g—qA, gPQf—, fAg—qA$ in Test I; $gPQ—, g—qA, fAgPQ—, g—qfA$ in Test II. The alternatives were again a $P$ word ($p_1$ or $p_2$), a $Q$ word ($q_1$ or $q_2$), and a word from one of the $A$ phrases. (The $PQ$ phrases correct in the short-sentence problems in Test I were correct in the long-sentence problems of Test II, and vice versa.) The problems were designed to be exactly parallel to Type 3 above, except that correct choices led to the construction of the four $PQ$ phrases to which $S$ had not been exposed ($p_1q_1, p_1q_2, p_2q_1, p_2q_2$). While the problems of Type 3 could be solved on the basis of rote learning of the $PQ$ phrases, Type 4 could presumably be solved only if $S$ had learned the internal structure of the $PQ$ phrase, i.e., that $p_1$ and $p_2$ go first after $g$, and $q_1$ and $q_2$ second.

The problems were presented in a random order in each test. The words (each on a separate card) were placed on the ledge of a board with the beginning and end of the sentence indicated, and with the space to be filled in specified by a gap. The alternatives were placed one above the other beneath the gap, the top-to-bottom placement of the correct alternative being chosen randomly. The $S$ was told to read over to himself the sentences that could be formed by inserting each alternative in the space, and to choose the alternative that made a sentence that he remembered hearing before. The $S$ was not informed of the correctness of his responses. At the beginning of the test he was told "I'm not going to tell you each time if you're right or not, but at the end I'll tell you if you got most of them right, and if you get most of them right, you win the rest of this pile of chips." The $S$ was always told he got most of them right.

After Test II, there was an unsystematically conducted "free-recall" test, whose purpose was to determine if learning had progressed to the point that $S$s could freely generate sentences. The words were laid out, and $S$ was invited to try to make sentences. He was stopped after he had made 10 error-free sentences, or earlier if he seemed unable to cope with this part of the task.

**Results**

Table 1 summarizes the results for the different types of problems. To compare the means with the values expected by chance (i.e., 2.0 for the two-choice problems, 1.33 for the three-choice), a $t$ test was used, basing the error term on whichever was larger, the sample standard deviation or the theoretical $\sigma (\sqrt{npq})$ computed from

---

<table>
<thead>
<tr>
<th>Test</th>
<th>Type</th>
<th>Problems Solvable</th>
<th>Mean Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1</td>
<td>20</td>
<td>2.00</td>
</tr>
<tr>
<td>I</td>
<td>2</td>
<td>20</td>
<td>2.00</td>
</tr>
<tr>
<td>I</td>
<td>3</td>
<td>20</td>
<td>2.00</td>
</tr>
<tr>
<td>I</td>
<td>4</td>
<td>20</td>
<td>2.00</td>
</tr>
<tr>
<td>II</td>
<td>1</td>
<td>20</td>
<td>2.00</td>
</tr>
<tr>
<td>II</td>
<td>2</td>
<td>20</td>
<td>2.00</td>
</tr>
<tr>
<td>II</td>
<td>3</td>
<td>20</td>
<td>2.00</td>
</tr>
<tr>
<td>II</td>
<td>4</td>
<td>20</td>
<td>2.00</td>
</tr>
</tbody>
</table>
ASSOCIATION AS A CONTEXT-SENSITIVE RULE

TABLE 1
MEANS AND SDs OF THE NUMBERS OF PROBLEMS CORRECTLY SOLVED OF EACH TYPE

<table>
<thead>
<tr>
<th>Test</th>
<th>Types of Problems</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2) short</td>
<td>(2) long</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>I</td>
<td>3.80 (0.52)</td>
<td>3.20 (0.62)</td>
<td>3.05 (0.89)</td>
<td>2.60 (1.27)</td>
<td>2.95 (1.05)</td>
</tr>
<tr>
<td>II</td>
<td>3.70 (0.57)</td>
<td>3.65 (0.67)</td>
<td>3.05 (1.05)</td>
<td>3.40 (0.50)</td>
<td>3.40 (0.50)</td>
</tr>
</tbody>
</table>

Note.—Four problems per cell. Three multiple-choice items used in Problem Types 3 and 4, two otherwise.

the binomial expansion. All the means were significant \( p < .001 \). (While this test assumes independence of trials within Ss, to the extent that there is nonindependence correlation would presumably be positive, and would render the test conservative.) The correct responses on the problems of Types 1 and 2 indicate that Ss knew how to form long sentences from short ones, and were able to recall which sequences followed which marker in both short and long sentence frames. The problems of Types 3 and 4 tested Ss’ knowledge of the composition of PQ phrases. Since the correct responses on Type 4 led to the construction of PQ phrases not previously experienced, they indicate that Ss had learned the internal structure of the PQ phrase; moreover, although Type 3 problems could have been solved on a rote-learning basis (since the PQ phrases constructed were the ones used in the exposure phase), it can be seen that these problems were not solved more frequently than Type 4, i.e., new PQ phrases appear to have been constructed about as readily as familiar ones.

Detailed comparisons among the various means in Table 1 seem inappropriate, because of the skewed distributions and heterogeneous variances due to differential ceiling effects, and because some of the problems are two-choice and some three-choice. Comparing Tests I and II globally, the total number of correct responses was greater on Test II \( p < .01 \), by \( t \) test; inspection of the table indicates that the improvement in performance occurred primarily on Problem Types 3 and 4. In both tests sex differences were small and not consistent as to direction on different types of problems.

Half the problems of Types 3 and 4 were of the form \( g—Q \), in which the space was contiguous with the marker element, and the other half were of the form \( gP— \), where the space was not contiguous with the marker. It might perhaps be thought that a contingency of a word on another word would be more readily learned if the items were contiguous. Combining Tests I and II, the mean numbers of correct responses were: 3.35 for Type 3 problems of the form \( g—Q \), 2.65 for Type 3 \( gp— \), 3.30 for Type 4 \( gP— \), and 3.05 for Type 4 \( gP— \) (four problems in each category). Each category was solved with significant frequency \( p < .001 \), indicating that correct responses occurred regardless of whether or not the item filled in was contiguous with the marker. In the Type 3 problems, significantly more problems of the form \( g—Q \) were solved than of the form \( gP— \) (by \( t \) test, \( p < .01 \)). However, in these problems, correct responses could reflect learned contingencies with the other P or Q words as well as with the marker. In the Type 4 problems, where correct responses can reflect a
contingency only on the marker, the difference was not significant \( (t = 1.16) \). These results do not suggest any great difference in the facility with which the positional definitions "second after \( g \)" and "first after \( g \)" are acquired.

A further analysis of the Problems 3 and 4 was made to see if errors on these problems showed any tendency for an A item to be chosen which occupied a position in the A phrase homologous to the vacant position in the PQ phrase. For the eight 3 and 4 problems in which such a choice was possible 13 Ss selected the A alternative on more problems than the wrong PQ phrase item and 4 Ss selected the wrong PQ phrase item more often \( (p < .05, \text{by sign test, omitting the 3 Ss who made both types of errors equally often}) \). On these problems the A alternative was actually selected on an average of 1.6 problems (as against .65 problems for the wrong PQ phrase item, and 5.75 problems answered correctly). In the other eight 3 and 4 problems, in which the A alternative was not homologously positioned, the A item was chosen only as frequently as the wrong P-phrase item (.7 problems each for the A- and the wrong PQ-phrase items, and 6.6 problems answered correctly). Comparing the frequency of choice of a wrong A-phrase item on the eight problems on which it was homologously positioned with the frequency on the eight problems in which it was not homologously positioned (i.e., 1.6 vs. 0.7), a homologously positioned A item was chosen on more problems \( (p < .01, \text{by } t \text{ test}) \). Thus, although errors were few, the favorite error was an item that occupied the same position in the other phrase.

In the "free recall," only two Ss made 10 correct sentences, and two others made 6 or more without making errors; it thus appears that the learning had not progressed to the point that Ss could freely generate sentences.

**Discussion**

The sentence-completion test results indicate substantial learning of the pattern properties of the exposed strings, including, in particular, the internal structure of the PQ phrase. The only basis for solving the Type 4 problems would appear to be a learning of the positions of \( p_1 \) and \( p_2 \) ("first after \( g \)") and \( q_1 \) and \( q_2 \) ("second after \( g \)"). An explanation of the responses on these problems in terms of mediation theory (Jenkins & Palermo, 1964) would appear ruled out because neither \( p_3 \) nor \( q_3 \) were ever exposed in combination with any of \( p_1, p_2, q_1, \) or \( q_2 \) so that some of the stages required in the stimulus-equivalence paradigm were not present (cf. Smith, 1965). Also a mediational explanation of the fact that the favorite error was a homologously positioned A-phrase item would obviously be very difficult to formulate. In general, therefore, this experiment provides evidence that Ss can learn positional relationships where position is defined relatively to a marker morpheme. The conditionality of the position learning on the marker may well be due to S simultaneously learning \( a \) the absolute position of the word in the phrase, and \( b \) contingencies between the words and the respective marker morphemes. The fact that the favorite error was a homologously positioned A item would then be due to somewhat better learning of \( a \) than \( b \).

The total time occupied by the two exposure periods together was of the order of 20-25 min. Although learning had not progressed so far as to allow free generation of sentences, nevertheless the Test II results indicate that a very substantial degree of learning had taken place during this relatively brief period. It appears that Ss are not merely able to learn the kinds of grammatical relations involved but learn them rather rapidly;
moreover, information feedback contrasting grammatical with ungrammatical strings is clearly not necessary for learning.

_Grammatical representation of what is learned._—Chomsky (1963) has described a hierarchy of types of grammars, which provide possible models of the kinds of regularities learned in experiments such as this. The simplest of these, a _finite state grammar,_ is a grammar all of whose rules are of the form $A \rightarrow xB$ or $A \rightarrow x$, where capital letters designate word or phrase classes (i.e., belong to the non-terminal vocabulary), and lower-case letters designate words (i.e., belong to the terminal vocabulary). Alternatively, and equivalently, the term refers to a state diagram (SD) of the usual type with no push-down storage annotations. (See Chomsky [1963] for details.)

In comparing possible grammars with experimental data, it is important to distinguish sharply between a grammar (a set of rules), and a language (a set of sentences): a given set of sentences may have several grammars, although, presumably, only one grammar will be "correct," in the sense that its rules represent the regularities actually learned by Ss. In the present experiment the sentences are a finite set. One possible grammar is therefore a simple list of the 85 strings. However, it is clear that Ss did not learn the sentences as a list of 85 different items, so such a grammar would not represent what was learned.

More generally, the regularities learned in this experiment, like those learned in a previous experiment (Braine, 1965b), do not appear to be describable in terms of the rules of a finite state grammar. The test results indicate that Ss learned that the components of the long sentences were the short sentences; thus a minimally adequate grammatical representation of the regularities learned must treat an $fA$ or a $gPQ$ sequence as being the same phrase (i.e., as having the same "P marker" or sequence of P markers), regardless of whether it occurred alone or as a part of a long sentence. Any finite state grammar with this property, and which generated the correct sequences $fA$, $gPQ$, $fAgPQ$, $gPQfA$, would necessarily incorporate a recursive rule which would generate also $fAgPQfA$, $fAgPQfAgPQ$, $fAgPQfAgPQfA$, . . . , and $gPQfAgPQ$, $gPQfAgPQfA$, . . . (note that this recursion is obtained if the push-down store is eliminated in SD-A in Fig. 1). No S in the free recall generated a string longer than six words, or any string which looked as if it might be an attempt to construct a sequence longer than two phrases; it was apparent that Ss had learned that the sentences were not more than two phrases long—one would surmise that this was one of the earliest properties they registered. Thus, a grammar which rendered the regularities learned would have to arrange that not more than one traverse through each phrase occurred. SD-A in Fig. 1 shows a device which appears to be a reasonable model of the relationships learned in all respects except one. It differs from a finite state diagram only in incorporating a small push-down store (PDS) which is used to control the entry into each phrase path, so that there is not more than one traverse through each. SD-A appears to be the minimally complex grammatical device (i.e., minimally more complex than a finite state diagram) which renders the principal regularities Ss learned.

The one aspect of the data which is not accounted for by SD-A is the finding that the favorite error in Problem Types 3 and 4 was a homologously positioned A item. This sort of error indicates a relation between the corresponding positions in the A and PQ phrases which has no counterpart in SD-A. SD-B in Fig. 1 shows a model which predicts this kind of error. In this model the A and PQ phrases are both generated through a single sequence of states, with homologically positioned items generated at the same state transitions. The contingencies among the words and the markers, which establish which word sequences belong to each phrase type, are indicated by means of the PDS entries. In effect, this model, relative to "A," achieves an
FIG. 1. Two possible PDS state diagrams rendering the relations learned. (Strings are
gen"erated" by movement from IN to OUT via the arrows joining states, cf. Chomsky, 1963,
pp. 339–345. A rule \([x, S_t, y] \rightarrow [S_j, z]\) is diagrammed as the triple \([x, y, s]\) between
\(S_t\) and \(S_j\). The first entry in a triple is the word that is generated at the transition, the
second entry is the item which must be on top of the PDS for the transition to take place,
and the third entry is the operation carried out on the PDS. Thus \([x, y, z]\) means "generate
\(x\) if \(y\) is the most recently stored item in the PDS, and store \(z\) on top of \(y\)."
\(e\) is the identity element: as first entry it means that nothing is generated, as second,
that the operation occurs regardless of what is in the PDS, and as third, that the PDS is left
unchanged. \(a\) means that the top PDS item is erased—bringing the item beneath it to the
top. S-PQ/PQ means that PQ in the PDS is erased and replaced by S-PQ; similarly
\(fa_1/f\) means that \(f\) in the PDS is replaced by \(fa_1\).)
economy of states by exploiting its PDS
to record features of structure. The
positions are represented by the transitions
between states, and the contingencies
between words by PDS entries. In
Model "B," any transitory forgetting of
the PDS entries, or any incompleteness
of learning of the connection between the
PDS entries and the words they control,
would lead to confusion as to which of
the words occurring at a state transition
should be chosen; confusion of this sort
should lead, in Problems 3 and 4, to a fre-
quent choice of the word occurring in
the homologous position of the other
phrase.

It is of considerable interest that if one
tries to represent the relations learned in
a grammatical notation (i.e., using a se-
quence of rewrite rules [Chomsky,
1963]), rather than by state diagrams,
the only apparent way of capturing the
difference between SD-A and SD-B is
by rendering SD-A by context-free (CF)
and SD-B by context-sensitive (CS)

rules. Thus, a CF grammar might con-
tain the rules: A phrase \(\rightarrow fA; A \rightarrow a, b, a_2, b_2, a_3, a_4; P phrase \rightarrow gPQ; P \rightarrow p, p_2, p_3, P \rightarrow q, q_2, q_3, Q \rightarrow q_1, q_2, q_3.
Like SD-A, such
a grammar provides no basis for predict-
ing the kinds of errors made. A CS
grammar might contain rules such as A
phrase \(\rightarrow fX; P phrase \rightarrow gX; X \rightarrow a_3, a_1\) in context \(f--; X \rightarrow YZ; Y \rightarrow a_1, a_2\) in context \(f--Z; Y \rightarrow P\) in context \(g--Z; P \rightarrow b_1\) in context \(fa_1--; Z \rightarrow b_2\) in context \(fa_2--; Z \rightarrow Q\) in context \(gP--; P \rightarrow p, p_2, p_3; Q \rightarrow q, q_2, q_3\). In this
grammar, the contingencies among the
items are represented in the contexts to
which the rules are sensitive; as in SD-B,
any temporary forgetting or incomplete
learning of the contextual sensitivities
leads to a merger of \(a_1, a_2,\) and \(P,\)
and of \(b_1, b_2,\) and \(Q,\) and thus to the
errors found.

Parenthetically, it seems to the writer
that the apparent correspondence of the
CF grammar to SD-A, and of the CS
grammar to SD-B, raises the question of
whether some strong equivalence theorem may not be provable between CF and CS grammars, and classes of automata differentiated according to the purpose to which their short-term stores are put. There appear to be two kinds of uses for a short-term store: to record each entry into a phrase path (with erasure of the record on exit), thus keeping a count on successive embeddings in nested constructions; and to indicate contingencies between elements of a string, as in SD-B. If the difference between these uses were made precise, a strong equivalence might be shown between CF grammars and PDS SDs which use their PDS only in the first manner. (Note that in Chomsky's [1963, pp. 371-373] construction converting a normal CF grammar into a PDS SD that is weakly equivalent, the PDS is used only to mark the beginnings and ends of phrases.) Similarly, CS rules seem to correspond to the second use of a short-term store: such a store would be "linear-bounded" (Chomsky, 1963, p. 338) rather than "push-down" when the contingencies in the CS rules straddle phrase boundaries.

These speculative correspondences between rewrite grammars and SDs aside, the difference between the A and B models above would appear to have interesting implications for the concept of an association in traditional S-R theory. The finite state model has been widely accepted as a model of "chains" of associations (e.g., Miller, 1951; Osgood & Sebeok, 1954, especially Sec. 5): in this model an association is represented by assigning the associated items to successive transitions in a sequence of states of an SD. Although SD-A above does go beyond a finite state system, it still represents the associations (i.e., the contingencies) among the words of each phrase in the manner of a finite state system. SD-B, however, represents associations in an entirely different manner: an association is a relation between an item and an entry in a short-term memory store. More precisely, an association between two items, a and b, is represented by a pair of rules \((a, S_p, x) \rightarrow (S_{i+1}, f)\) and \((b, S_p, f) \rightarrow (S_{i+1}, y)\) where \(S_j\) is a state subsequent to \(S_i\) in a left-to-right pass through the diagram, and \(f\) may be reasonably identified with either a or b. \((x\) and \(y\) are irrelevant, and if a and b are associated only with each other, then \(x = e\) and \(y = e).\) Alternatively, and more speculatively (cf. preceding paragraph), an association might be viewed as a context-sensitive rule. These models would provide a much more powerful and flexible concept of an association than the finite state model.

Although the conclusion that at least a PDS model of associations is required is based on a relatively subtle feature of the data in this experiment, the conclusion agrees closely with the results of a previous experiment (Braine, 1965b). In that experiment, Ss were exposed to a set of sentences, half of which were of the form \(aXb\) and half of the form \(pq\), where \(a, b, p, q\) were individual words and X was a word from a list of 18 items. Most Ss readily learned the \(a-b\) and \(p-q\) contingencies, and the X items were learned, not as two lists, one associated with \(a . . . b\), and the other with \(p . . . q\), but as a single list which went second, or "in the middle." The regularities learned cannot be adequately represented by a finite state grammar, for several reasons (Braine, 1965a, 1965b; Gough & Segal, 1965). At least a phrase structure model is required, or a PDS SD in which the \(a-b\) and \(p-q\) associations are represented in the manner outlined above. Both this experiment and the present one have in common the fact that they involved the formation of "remote" associations between items whose positions were also learned. Remote associations are, by definition, associations which straddle intervening material. It seems intuitively clear that such associations are poorly translated in finite state diagrams where nothing corresponds to the "straddling" property. It may be that the tacit acceptance of the finite state model as adequate for representing learned associations is due only to the fact that the actual properties of remote associa-
tions have been relatively little explored in the verbal learning literature.

REFERENCES


(Received July 30, 1965)