


For the psychologist, the interesting suggestion made by the concept of phrase-structure rules in language is the possibility that they may represent a formally stated analogue of the psychological rules used for encoding language. That is, the string of morphemes or words encompassed by a single phrase-structure rule may represent a psychological unit which is determined by the Ss encoding rules.

For an S, language material represents a sequence of coordinated movements of chest, larynx, tongue, mouth, lip, etc. It is reasonable to assume that it is not necessary for him to learn how to produce the motor responses which result in the sounds. In general, these are responses which previously have been integrated into units. Further, S should not have to integrate or learn the sequence of responses that result in culturally frequent words (e.g., *boy, girl, the, as, etc.*). Again, these sequences should have been integrated into units as a result of previous language experience.

While it is reasonable to assume that the sound elements (phonemes) and at least certain high-frequency words are integrated units, it also may be the case that certain recurring sequences of word classes (e.g., phrases) represent integrated response units. That is, when Ss are presented the task of learning grammatically structured verbal material, it may be that part of the learning has already occurred.

These considerations suppose that there are item pools which consist of response sequences that have been integrated, or learned as sequences, and hence can be sampled as relatively simple response units in new learning situations. That is, in a new learning task the response-integration stage illustrated by McGuire (1961) need not occur because it occurred during previous learning experiences.

The clearest evidence of the phenomenon in a simple paired-associate task is an S’s tendency to emit an integrated response either perfectly or not at all. The degree to which

---. The research reported herein was supported through The Cooperative Research Program of the Office of Education, U.S. Department of Health, Education, and Welfare.
a response sequence has been integrated into a unit should be inversely proportional to the probability of giving a response which is partially correct.

When partial sentence responses are emitted during a learning task they should begin and end at the boundaries of encoding units. For example, if S is required to learn sentences, then, during learning, the transitions from correct words to either incorrect words or no word should conform to the boundaries between units of encoding. If each of S's response attempts is scored for transitional errors (TE) by determining the conditional probability of each word in the sentence being incorrect, given the preceding word is correct (i.e., the probability of a transition from a correct word to an error on the next word) then, if Ss are encoding in subsentence units, there should be jumps in the probability of a transitional error (TE) at the beginning and end of a unit.

The next step is to determine the degree to which these between-boundary word strings conform to sequences generated by phrase-structure rules. At the onset of the present study it was assumed that Ss would select some particular order of phrase-structure rule as the unit into which they encode their language. As a first approximation this was done by calculating the frequency of each word in each phrase in the entire text. The phrase units used were simple noun, verb, prepositional, and parenthetical phrases.

The following experiment was intended to test the hypothesis that there are word-to-word transitions within sentences where the probability of a transitional error (TE) is significantly greater than for other transitions. More specifically, these points should occur at transitions from one phrase to the next. For the sentences in Table 1 this point should be between the third and fourth word for List 1 sentences, and between the second and third, and between the fifth and sixth words for the List 2 sentences.

**Method**

**Materials**

The learning materials used were the sentences listed in Table 1. There were 16 sentences, eight of them with two major phrases (List 1) and 8 with three major phrases (List 2). All sentences were composed of seven Thorndike-Lorge A and AA words, and the total number of letters in the List 1 sentences is equal to the total number in the List 2 sentences. Figure 1 gives a tree diagram for an example sentence from each list. The sentences were constructed by giving the tree diagrams to a linguist naive to the purpose of the experiment and asking him to make up eight sentences that fit each diagram using only A and AA words, and, as much as possible, not repeating any words.

**Procedure**

The Ss' task was to learn an eight-item paired-associate list, in which the digits from 1 to 8 were the stimuli and the sentences in Table 1 were the responses. Group 1 learned the two-phrase sentences in List 1 and Group 2 learned the three-phrase sentences in List 2.

The learning materials were presented on a Lafayette memory drum at a 4:4-sec rate, with a 4-sec intertrial interval. There were four randomizations of the PA list and all Ss received a constant 13 trials.
and between the fifth and sixth List 2 sentences.

METHOD

The materials used were the sentences. There were 16 sentences, eight of major phrases (List 1) and eight with minor phrases (List 2). All sentences were composed of Hornikke-Lorge A and AA words. The number of letters in the List 1 sentence total number in the List 2 sentence gives a tree diagram for each sentence. The sentences were grouped into two lists.

Table 1

<table>
<thead>
<tr>
<th>List 1</th>
<th>List 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>saved the dying woman.</td>
<td>across the street is burning.</td>
</tr>
<tr>
<td>people gave a large party.</td>
<td>who left early are listening.</td>
</tr>
<tr>
<td>farmer likes the old horse.</td>
<td>with black tails are running.</td>
</tr>
<tr>
<td>work tires many young men.</td>
<td>who sings well was found.</td>
</tr>
<tr>
<td>child picked a blue flower.</td>
<td>on the lake are cooking.</td>
</tr>
<tr>
<td>man ate the fresh bread.</td>
<td>that I heard was singing.</td>
</tr>
<tr>
<td>man found the wrong number.</td>
<td>with red bows were sold.</td>
</tr>
<tr>
<td>lady broke the green dishes.</td>
<td>who live here are dancing.</td>
</tr>
</tbody>
</table>

![Tree diagrams](image)

FIG. 1. Tree diagrams of the sentences. The top diagram represents the List 1 sentences and the bottom the List 2 sentences. The diagram is determined by the immediate constituent structure of the sentence and is a way of graphically representing the structure. The top circle or node stands for the entire sentence as a single construction and the two lines leading from it indicate the constituents into which it is divided, etc.

The Ss were given instructions which emphasized that they should give as much of each response as they could remember, even though they could not remember all of it. In addition, if the Ss were not attempting to guess during the first few trials E encouraged them to do so.

Subjects

The Ss were 36 introductory psychology students who participated as part of their course requirement. Half of the Ss learned one list and half the other list, and they were assigned to the lists in alternating order as they appeared for the experiment.

Scoring

Probability of a Transitional Error (TE). The probability of a TE within a sequence was computed by counting the frequency with which a particular word was wrong given the preceding word was correct and dividing that frequency by the total frequency that the preceding word was correct (i.e., dividing the TE frequency by the number of opportunities for a TE). By computing the probability in this manner a correction is made for any tendency the Ss might have to integrate the response in serial order.

Conditional Probabilities of Phrase Units. It follows from the hypothesis outlined above that if an S says at least one word within a phrase correctly, the probability of another word being correct is considerably greater than would be predicted if each word was an independent event. Under the assumption that the occurrence of words within a sequence is independent events, the maximum conditional probability of saying all n words correctly, given that at least one was correct, is the mean probability of saying any one word raised to the power of n-1. For a three-word phrase the prediction as to the conditional probability of emitting the whole phrase correctly, if independence is assumed, is the square of the mean probability of emitting any one word correctly. For a two-word phrase the prediction is simply the mean probability of emitting any one word correctly. The empirical probabilities were determined by dividing the total frequency with which a phrase was emitted perfectly by the frequency that at least one word was emitted correctly (i.e., what proportion of the times that something was correct, was all of the phrase correct.).

RESULTS

Sentence Learning Rates

To test the difference between the groups each S was scored for the total number of words given correctly across the 13 trials. The mean for Group 1 is 337, and for Group 2 it is 365, t (54) = .82, p > .40. It appears, then, that while the linguistic form of the responses is different for the two groups, the rates of learning these responses did not differ significantly.

Each of the two forms of sentence used in the experiment can be viewed as having seven word-positions. To determine if there was any tendency for the Ss to integrate the sentences in sequence from the first word to last, each S was scored for the total frequency over 13 trials with which he emitted any of the first three words correctly as opposed to the frequency for the last three words. For both groups the first part of the sentences was emitted correctly more often than the last part, for Group 1, t (27) = 1.89, p < .10; for Group 2, t (27) = 5.9, p < .001. In both cases, then, there was a tendency for the Ss to integrate the sentence responses as a serial task from first word to last.
Pattern of Transitional Error Probabilities

The probability of a TE for each of the transitions for each of the two forms of sentence is presented in Table 2. For each S the probability of a TE was computed for each of the six transitions by using all eight sentences over the 13 trials. For each group the data were arranged in a $6 \times 28$ table and a Friedman two-way analysis of variance was performed. The question to be answered by the analysis of variance is whether the probabilities of a TE within a response tend to be equal or whether they are different, as would be the case if the Ss learned the sentences in units which were larger than a word but shorter than a sentence. The $\chi^2$ (5) for Group 1 is 62.01, $p < .001$, and for Group 2 it is 62.90, $p < .001$. For both groups the six probabilities were significantly different from one another.

If phrase structure is used as a basis for predicting TE probability then the Group 1 Ss should tend to have a greater probability of a TE on transition 3 than on the other transitions and the Group 2 Ss should tend to have a higher probability of a TE on transitions 2 and 5. Each S was given two scores, the mean probability of a TE on the predicted transition or transitions and the mean probability of a TE on other transitions, and the differences were tested by means of a Wilcoxon matched-pairs signed-ranks test. For Group 1 the respective means are .120 and .059 ($z = 3.14$, $p = .0008$), and for Group 2 they are .129 and .026 ($z = 4.62$, $p < .00003$). The Ss did tend to have unequal probabilities of a TE within a response, and the high points tended to be on the transitions predicted from the phrase structure of the sentences.

For both phrases in the List 1 sentences and the middle phrase of the List 2 sentences there is more than one intra-phrase transition. The original prediction was that these intra-phrase TE probabilities should be equal if the phrase is the response unit the Ss were using. The Wilcoxon test was used, and for the List 2 phrase $z = 1.50$ ($p = .1336$), whereas for the first List 1 phrase $z = 2.34$ ($p = .0192$). A Friedman two-way analysis of variance was used for the second List 1 phrase because it had three transitions, and $\chi^2 (2) = 84.9$, $p < .001$. Contrary to the original hypothesis, the intra-phrase TE probabilities seem to be unequal at least for the List 1 sentences.

An inspection of the diagrams of the sentences in Fig. 1 suggests an interesting relation between the diagrams and both the intra-phrase and inter-phrase pattern of TE probabilities for the respective sentences given in Table 2. The diagrams can be viewed as reflecting the constituent structures of the sentences. Each of the nodes (i.e., circles) in the diagrams represent a construction, and the lines leading from a node indicate the two constituents into which it is to be divided. In a sense, then, the node represents the point in the linguistic analysis of the sentence where a division is made between the last word of the first constituent and the first word of the last constituent. At some point in the analysis every word is divided from its adjoining words by nodes, and, within limits,

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>Transition No. 1</th>
<th>Transition No. 2</th>
<th>Transition No. 3</th>
<th>Transition No. 4</th>
<th>Transition No. 5</th>
<th>Transition No. 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two-phrase</td>
<td>.108</td>
<td>.049</td>
<td>.120</td>
<td>.070</td>
<td>.032</td>
<td>.020</td>
</tr>
<tr>
<td>Three-phrase</td>
<td>.015</td>
<td>.143</td>
<td>.073</td>
<td>.020</td>
<td>.118</td>
<td>.012</td>
</tr>
</tbody>
</table>
phrases in the List 1 sentences than one intra-phrase transitional prediction was that these TE probabilities should be equal to the response unit the Ss were given. The response unit the Ss were given for Ss at this phrase is 1.50 (.p = .1336), and for the first List 1 phrase 1.50 (.p = .1336), and for the first List 1 phrase 2.34 (.p = .1336) (2.34 (.p = .1336). A Friedman two-way analysis was used for the second List 1 phrase. It had three transitions, and the probability of a TE was .75 (.p = .1336). Contrary to the hypothesis, the intra-phrase TE probability was unequal at least for the Ss.

Of the diagrams of the sentences in this experiment, an interesting relation between the intra-phrase and inter-phrase TE probabilities for the respective sentences is found. The diagrams can be viewed as constituent structures of the sentence nodes (i.e., circles) in the diagram construction, and the lines leading to the two constituents into which the sentence is divided. In a sense, then, the node representation is the linguistic analysis of the division made between the last constituent and the first word of the sentence. At some point in the analysis, the words of the sentence are divided into its adjoining words by the simple limits, it seems possible to rank the features of the "level" at which the transition occurs. For example, for the first division is between the first and second words at transition 5; the second division is between the third and fourth words at transition 2; the third division is at transition 2, the fourth division is at transition 3, the fifth division is at transition 4, and the sixth division is at transition 5. The same kind of analysis can be made for the list 1 sentence

**REALITY OF PHRASE-STRUCTURE RULES**

For purposes of a check on the reliability of these results, data were available from a pilot study which was identical to the present experiment, except that one group of 30 Ss learned a list composed of the first four List 1 sentences and the first four List 2 sentences while the other group of 30 Ss learned a list composed of the other eight sentences (i.e., the lists were heterogeneous with respect to the grammatical form of the responses).

The Kendall rank-correlation between the pattern of TE probabilities obtained in the two experiments is .60 (.p = .068) for the List 1 sentences, and .73 (.p = .018) for the List 2 sentences. These correlations suggest that while the differences in probability of a TE at the various transitions are small, they are quite reliable.

For the pilot study data the correlations (tau) between level at which linguistic divisions occur and the probability of a TE is .50 (.p = .136) for the List 1 sentences, and .75 (.p = .068) for the List 2 sentences. While these values are smaller than those obtained in the present experiment, they do tend to substantiate the possibility of predicting the probability of a TE from the phrase-structure of a sentence.

**Phrase Conditional Probabilities**

A trial-by-trial plot of the phrase conditional probabilities represent a learning curve for the rate at which the phrases were integrated. If one examines these obtained curves in relation to the predicted curves generated on the assumption of independence, the Ss' initial tendency to treat the phrases as integrated units can be seen, as well as the increasing integration which occurs during the learning task itself. Examples of obtained and predicted curves are presented in Figs. 2 and 3. In Fig. 3 the theoretical conditional probability of saying a whole phrase is simply the conditional probability of saying one word because it is a two-word phrase. These figures

<table>
<thead>
<tr>
<th>SENTENCE FORM</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2</td>
<td>.70</td>
<td>.012</td>
<td>.020</td>
<td>.108</td>
<td>.018</td>
<td>.020</td>
</tr>
<tr>
<td>3 4</td>
<td>.70</td>
<td>.012</td>
<td>.020</td>
<td>.108</td>
<td>.018</td>
<td>.020</td>
</tr>
<tr>
<td>5 6</td>
<td>.70</td>
<td>.012</td>
<td>.020</td>
<td>.108</td>
<td>.018</td>
<td>.020</td>
</tr>
</tbody>
</table>

**Fig. 2.** Conditional phrase probabilities for the first phrase of the List 1 sentences. The probabilities were computed across all sentences and all Ss. Initial

**Fig. 3.** Conditional phrase probabilities for the first phrase of the List 2 sentences.
represent the first phrase of each of the two sentence types, and the other three curves look about the same.

Each S was given two scores for each phrase. One score was the mean probability of any one word being correct and the other was the conditional probability of the whole phrase being correct, given something was correct. In all five cases the mean probability, across all trials, of emitting the whole phrase correctly exceeds, beyond the .01 level, that which would be predicted under the assumption of independence. An inspection of the figures indicates that the differences between the predicted and obtained curves were not only the result of a gradual integration during the learning experience but the fact that they also occurred on the very first test trials suggests the Ss approached the task with a marked initial tendency to treat the phrases as units.

**DISCUSSION**

The results of this experiment are generally consistent with the hypothesis that Ss approach language material in terms of pre-integrated units, and that the units are predictable from the linguistic structure of the material, (i.e., to a degree, at least, phrase-structure rules seem to be "psychologically real").

The most interesting result, however, is the relation between the probability of a TE and the linguistic "level" of the transition. Originally it was expected that Ss would tend to learn the sentences in terms of a single order of unit which would be roughly equivalent to phrases. That expectation led to the prediction that the probability of a TE on intra-phrase transitions should be relatively low, and approximately equal, while the inter-phrase transitions should show relatively high probabilities of a TE. The results support the hypothesis to the extent that (a) the probability of a TE is higher for inter-phrase transitions than intra-phrase transitions; and (b) there is a higher conditional probability of the whole phrase occurring, given something was said, than would be predicted from the assumption that each word was an independent event.

While those results do support the idea that there may be a functional unit conforming to the phrase, they do not necessarily imply it is the only unit. The probability of an isolated word suggests that there may be a within-phrase structure as well as a between-phrase structure.

The suggestion that language substructures may themselves be substructured has not only been made by linguists, but psychologists as well (e.g., see Braine, 1963; Miller, Galanter, and Pribram, 1960; Osgood and Sebeok, 1954). Miller et al. have expanded the idea further in an attempt to describe all sequentially ordered behavior in terms of a hierarchy of response units, with higher level units encompassing the lower level units. They have suggested that, on a behavioral level at least, rather than conceive of language sequences as sets of serial dependencies, one can view them as sets of "nested dependencies" reflecting the units-within-units concept.

If one is to use a model which assumes some kind of serial dependency or association between adjacent response items to explain the results of the present study, it would be necessary to suppose that the unequal TE probabilities are the result of unequal pre-existing habit strengths between the words. To explore that possibility, a set of serial and paired-associate lists were constructed from the List 2 sentences in order to study the differential learning rates of transitions 2 vs. 3 and 3 vs. 6. (e.g., house-across and street-is vs. across-the and is-burning). Twenty Ss learned each List for 15 trials and they were scored for number of errors during the 13 trials. When these transitions were included within the List 2 sentences the second and fifth transitions showed a greater probability of a TE. However, when the transitions were taken out of the context of the sentence, the two kinds of pairs were learned at the same rate under both serial and paired-associate conditions. What small differences that did occur were in favor of transitions 2 and 5.

For the sentences used in the present experiment a schema similar to that given in the diagrams of Fig. 1 might be adopted as an alternate to a simple serial dependency model. The correlations reported
REALITY OF PHRASE-STRUCTURE RULES

above offer some empirical support for such a decision. If the nodes (circles) in the diagrams are assumed to represent successively more inclusive encoding units, then generation of the sentences can be viewed as proceeding through a series of decoding steps rather than executing a sequence of word-to-word habits. Any particular encoding unit can be decoded into (a) response items (e.g., words or morphemes), (b) other encoding units, or (c) response items and encoding units. When an encoding unit has been generated it is decoded into its constituents if $S$ knows the decoding rules.

A reasonable approximation to the TE data can be obtained by stipulating that the probability that an encoding unit will be decoded is some function of the number of further decoding operations needed before the unit results in response items. Such a proposition not only relates to the "linguistic level" idea mentioned above, but also carries a testable hypothesis (albeit ad hoc) regarding why there should be such a correlation between "level" and TE probability.

REFERENCES


Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychol. Rev., 1956, 63, 81-97.


(Received August 10, 1964)