Grammar and the Recall of Chains of Verbal Responses

KURT SALZINGER
Biometrics Research, New York State Department of Mental Hygiene and Polytechnic Institute of Brooklyn

AND

CAROL ECKERMAN
Columbia University, New York 10032

The validity of transformation grammar for explaining differences in recall of differently structured series (a simple declarative sentence—D vs. a passive negative query—PNQ) of 10 nonsense syllables and function words was tested. All the main variables, except for grammatical structure, gave rise to significant differences. Function words were more easily recalled than nonsense content words; sentences than random arrays; whole presentation was superior to serial presentation; and the second sentence was recalled better than the first. Interactions showed that differences between sentence and random order, between D and PNQ, and between whole and serial learning were markedly reduced from first to second presentation. Function words depended more on contextual constraint than nonsense syllables, and D and PNQ differed most when presented serially in random order. The results could not be explained by a transformation grammar model but were consonant with the concept of frequency of occurrence of single words and combinations of words.

Chomsky (1957) has claimed that an approach to grammar in terms of frequency of single words and frequency of their combinations is inadequate, and he has promulgated a transformation grammar which psychologists such as Miller (1962) have used to investigate the psychological aspects of grammar.

The validity of the transformation model of grammar—namely, that there is a set of rules which allows the S to change or transform the simplest kind of sentence, the kernel, into the more complicated structures—is, of course, an empirical problem. The simple declarative sentence in the active mood—"The boy threw the ball"—would be an example of a kernel,

2 A recent article by Bever, Fodor, and Weksel (1965) states that psychologists like Miller (1962) and Mehler (1963) have made the mistake of identifying the kernel with the simple declarative sentence and have assumed that transformations are performed upon the kernel sentence when in fact the correct interpretation should have been that the simple declarative sentence is only one of the types of sentences which can be generated by a basic kernel structure. Whether one looks upon the kernel as a sentence or a "structure," the transformation grammar of Chomsky (1957) does suggest a difference in the number of steps necessary to generate the two sentence types discussed in this experiment and this is the concern of this study.
and the passive negative query—"Wasn't the ball thrown by the boy?"—an example of a complicated structure. Miller (1962) and Mehler (1963) suggest that memory for complex structures depends on the coding of the kernel sentence plus some indication of what transformation must be performed to rearrange the words into their correct and more complex order. It follows from this that when Ss are given kernel sentences to memorize, they should make fewer errors than when given more complicated structures, and this is in fact reported (Mehler, 1963). Unfortunately, the prompting method of recall used by Mehler, which consists of supplying S with discriminative stimuli for the lexical words and not for the function words which serve to differentiate the structures of the sentences, would be most helpful to kernel sentences and might therefore have produced an artifact in his experiment. Furthermore, the different number of words in the various structures—from 6 for the kernel to 8 in the passive negative query (PNQ)—should have been controlled to assure that the smaller number of words alone, or in interaction with the prompting procedure or in interaction with the differences in structure, could not explain the differences in memory between the kernel and the other sentences.

In view of these somewhat equivocal results and in order to test the theory, it seemed reasonable to investigate the degree to which sentences of different grammatical structures can be recalled. Since Chomsky (1957) has claimed independence of grammar and meaning, it was decided to use a technique recently employed by Epstein (1961), namely, nonsense syllables with bound morphemes and function words so arranged as to provide the discriminative stimuli for different sentence structures. Two structures were employed which were described as most different in Miller's (1962) paper, i.e., the simple active declarative sentence vs. the passive negative declarative sentence. Furthermore, the same number of word units was employed in both structures to eliminate the possible artifact of Mehler's experiment. Finally, by using a condition in which the units were presented in nonsentence order it was possible to measure the influence of grammatical structure by itself on memory and thus to replicate and extend (by including an additional structure) the experiment first done by Epstein (1961).

**METHOD**

**Materials.** Two grammatical types of 10-unit "sentences," each consisting of two different sets of function words and nonsense syllables, were constructed. One grammatical structure was a simple declarative sentence (D) and the other a passive negative query (PNQ). The nonsense syllables selected for the experiment had association values ranging from 73 to 89% (Glaze, 1928); disyllables used varied between 1.26 and 1.28 in meaningfulness (Noble, 1952). The required characteristics of English syntax were simulated by adding the prefix "be" and "de" to indicate verbs, and the suffixes "s" for nouns, "y" and "er" for adjectives, "ly" for adverbs, and "ed" and "ing" for verbs, as well as by using function words. The frequency of occurrence of successive bigrams and trigrams of the disyllables (Underwood and Schulz, 1960) was approximately equated for the two sentence sets (a and b in Table 1). Inspection of Table 1 shows that the same two basic sets of units were employed in all the word series. However, the following exceptions were made to keep the number of word units the same for the two grammatical sentences: the function word "and" was used in the D sentence to match the required function word "by" in the PNQ sentence, the verb ending "ing" in D was matched by the ending "ed," and the word "were" in D was matched by the word "weren't" in PNQ. The first word in each of the four sentences was capitalized; the two D sentences ended with a period; and the two PNQ sentences ended with a question mark. The order of the words in each of the sentences was randomized according to the same random order to produce four unstructured series. Both capitalization and punctuation were omitted from these random series.
TABLE 1

THE 10-WORD STIMULUS SERIES

<table>
<thead>
<tr>
<th>Declarative grammatical</th>
<th>Set a—And the piqy kews were beboving the nazer zumaps dygly.</th>
<th>Set b—And the kavy bycs were derizing the tober latuks neply.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Declarative random</td>
<td>Set a—beboving piqy nazer were the and zumaps kews dygly the</td>
<td>Set b—derizing kavy tober were the and latuks bycs neply the</td>
</tr>
<tr>
<td>Passive-negative-query grammatical</td>
<td>Set a—Weren’t the nazer zumaps dygly beboved by the piqy kews?</td>
<td>Set b—Weren’t the tober latuks neply derized by the kavy bycs?</td>
</tr>
<tr>
<td>Passive-negative-query random</td>
<td>Set a—beboved nazer the dygly by weren’t piqy zumaps kews the</td>
<td>Set b—derized tober the neply by weren’t kavy latuks bycs the</td>
</tr>
</tbody>
</table>

The verbal material was prepared for presentation on a Lafayette memory drum (Model 303) by both a serial and a whole method. For the whole method of presentation, each series was typed in a single horizontal line across a separate 8.5 × 11-inch sheet of white paper. The series was centered on the page and one space separated successive words. For the serial method of presentation, the words of a given series were typed successively in the middle of the page with the second word centered under the first and so on. Two lines were skipped between successive words and any punctuation immediately followed the last word.

A random series of ten single digits was constructed with the aid of a table of random numbers, with the restrictions that all digits from 1 to 9 inclusive be used and that the same digit appear in both the second and seventh positions to follow the repetitions occurring in the D sentences. This series of digits was also prepared for both the serial and whole methods of presentation. Four spaces occurred between successive digits in the whole method.

Subjects. Eighty undergraduate and graduate students whose native language was English were employed. Their ages ranged from 18 to 28 years.

Procedure. The Ss were tested individually and paid $1.50 an hour for participating. They were assigned to one of 16 experimental conditions which consisted of all combinations of the variables listed below, except that a given S received both sets (a and b) of stimulus material under the exact same conditions. The 16 conditions were permuted on the basis of random orderings of the numbers 1 through 16. The Ss were then placed into each of the conditions in the order in which they volunteered for the experiment. When one set of 16 conditions was completed, Ss were then placed, in the order in which they volunteered, into the next set of 16 randomly arranged conditions. This process was continued until data had been collected from 80 Ss. The variables combined were: grammatical structure (D vs. PNQ); word order (sentence order vs. random arrangement of words); stage of learning (first vs. second presentation of a or b); method of stimulus presentation [entire series at once (whole) vs. one unit at a time (serial)]. For purposes of analysis, errors were counted separately for nonsense syllables and for function words, thus defining a fifth main factor used in the analysis of variance. All Ss were presented with the series of numbers first: the entire series (for 10 secs.) or one unit at a time in agreement with the presentation conditions of the material which followed.

For the whole method of presentation, Ss were told to learn the numbers in the order in which they were presented and to distribute their attention equally among them. After each stimulus presentation they had 30 secs. to write the numbers they recalled on a sheet of paper provided with ten blanks.

The Ss in the serial presentation condition were given substantially the same instructions except that they were told the numbers would appear successively in the window of the machine. The single digits appeared at a rate of 1 per sec.

In both the serial and whole methods the procedure was repeated until S wrote the digits correctly on three successive trials. Immediately after a trial, E checked the record. To be correct, all the digits had to appear in correct order. The
S was given no knowledge of results until after the third successive correct trial, at which point E said, "good" and asked if S had any questions about the procedure. When it was ascertained that S understood the procedure, E proceeded to the presentation of the first series of nonsense and regular words.

The Ss receiving the grammatically structured series (D or PNQ) by the whole method were told that they would be shown a sentence composed of both nonsense and regular words for 10 secs. and were then given the same basic instructions, with appropriate changes, as for the numbers.

The Ss administered the same grammatically structured series as above by the serial method were given substantially the same instructions except that they were told that the words would appear successively in the window.

The Ss given the random series (whole or serial) received the same basic instructions but were told that E was presenting words originally arranged to make up a sentence, but which were now scrambled so that they no longer read like a sentence. The Ss were still required to learn the words in the order in which they were presented.

Words were counted as “correct” when they were written in correct order and spelled the same as the stimuli. For three Ss who made the same single spelling error for 10 successive trials, the experiment was terminated at that point, and the data included in this study. One S had the PNQ sentence (Set a) presented serially, one the D sentence (Set b) presented serially, and the third had the D sentence (Set b) presented whole.

After each S learned the first word series, E said, "good" and immediately proceeded to the second word series following the same procedure.

**RESULTS AND DISCUSSION**

Recall of Numbers. Statistical analysis showed that it took Ss almost twice as many trials to reach a criterion of one 10-unit sequence correctly recalled when using the serial method ($M = 4.15$) than when using the whole method ($M = 2.13$), $t(78) = 5.19, p < .001$. This difference remained for comparison of whole vs. serial on a criterion of two (5.72 vs. 3.27, respectively) and three successively correct recalls (6.72 vs. 4.35, respectively), $t(78) = 5.52, p < .001$. It indicates that presentation of an entire sequence of units rather than one item at a time will result in faster learning even when the sequence is not arranged according to some underlying order. This difference is probably due to the fact that S can distribute his attention so that combinations more difficult to learn get more practice. Furthermore, S can arrange the items in larger units when all are exposed at once.

Recall of Nonsense Syllables and Function Words. The recall data were converted into two basic scores: number of errors per function word and number of errors per nonsense syllable. A five-way analysis of variance was then performed on these basic scores. Table 2 shows the mean error scores for all the conditions.

<table>
<thead>
<tr>
<th>Stage</th>
<th>D Series</th>
<th>PNQ Series</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sentence</td>
<td>Random</td>
</tr>
<tr>
<td></td>
<td>Whole</td>
<td>Serial</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First</td>
<td>1.28</td>
<td>2.53</td>
</tr>
<tr>
<td>Second</td>
<td>1.57</td>
<td>2.33</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stage</th>
<th>Nonsense syllables</th>
<th>Function words</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First</td>
<td>1.28 .53</td>
<td>.18 .55</td>
</tr>
<tr>
<td>Second</td>
<td>1.60 .10</td>
<td>.28 .75</td>
</tr>
</tbody>
</table>
for the criterion of 3 successive correct recalls. Analysis of variance to a criterion of 1 correct recall shows substantially the same results, the exception being that only for the lower criterion is there a significantly greater number of errors for the PNQ than for the D structure, \( F(1, 288) = 3.93, p < .05 \).

Error scores were used in preference to trials to criterion for this analysis in order to include the difference in recall between function and nonsense words within the same analysis as the other variables and, perhaps more important, in order to evaluate interactions of word type with the other variables. Analysis of variance applied to trials to criterion scores showed substantially the same results as the error scores analysis, i.e., no significant difference due to the grammatical structure variable, \( F(1, 128) < 1, p > .05 \); a significantly large number of trials to criterion for the first than for the second learning stage, \( F(1, 128) = 7.65, p < .01 \); and nonsignificant trends in the same direction as the error score analysis for the word order variable (sentence vs. random) and the method of stimulus presentation variable (whole vs. serial).

**Grammatical Structure.** The difference in memory due to grammatical structure is at best equivocal, \( F(1, 288) = 3.07, p > .05 \). The Grammatical Structure-Stage of Learning interaction, \( F(1, 288) = 9.71, p < .005 \), shows a larger number of errors in the PNQ structure than in the D structure for the first presentation only, i.e., at the beginning of learning only. These results suggest that earlier relative unfamiliarity with the PNQ structure is simply overcome by presenting S with that structure. In terms of learning theory, the successive words of the PNQ sentences may be posited to evoke other than the called for words at the beginning of learning because of the higher frequency of occurrence of these competing responses. If Ss did in fact commit the kernel and the transformation rule to memory separately, then one should find more errors for the PNQ than for the D sentence for both presentations.

Accepting the equivocal difference between grammatical structures as real, further inspection of the conditions under which grammatical structure affects recall demonstrates that PNQ gives rise to more errors than D only in serial presentation, i.e., the interaction is significant, \( F(1, 288) = 5.10, p < .05 \). Following Epstein's reasoning (1962), we should expect such a difference only for whole presentation, since only this would allow S to learn the underlying structure. The triple interaction between grammatical structure, method of stimulus presentation, and word order shows still another reason for rejecting the notion of a plan or underlying structure, \( F(1, 288) = 5.42, p < .025 \), since the difference between D and PNQ stems almost wholly from the difference of the random arrangements of D and PNQ under serial conditions of presentation. The reversal under whole conditions of presentation clarifies the interaction between grammatical structure and method of presentation; apparently, whole presentation allows S so to distribute his attention over the items as to compensate for the differences in difficulty. It would seem most reasonable to conclude that the grammatical structures do not produce differences as structures but rather in terms of the frequency of occurrence of their words or combinations of words. The fact that the words of the PNQ structure produce a large number of errors when in random arrangement raises the possibility that they are generally combined in fewer different ways than the words of the D structure. In other words, the random arrangement of the PNQ structure results in a relatively greater
number of chunks (Miller, 1956) or units (Salzinger, 1962) than the randomly arranged D words, since the words in the PNQ sentences are less frequently used in other combinations. It is certainly most difficult to explain this result in terms of transformation grammar, which presumably is irrelevant to random (ungrammatical) sequences.

**Word Order.** These results are in agreement with Epstein's findings (1961, 1962), that words in random order produce a greater number of errors than the same words in sentence form, \( F(1, 288) = 18.61, p < .005 \). It is interesting to note, however, that here, as for the grammatical structure variable, the difference due to word order is evident only early in learning, viz., word order-stage of learning interaction, \( F(1, 288) = 9.57, p < .005 \).

The triple interaction of word type, method of stimulus presentation, and word order indicates that the difference in recall due to word order does not manifest itself at all for nonsense syllables when they are presented serially, \( F(1, 288) = 5.35, p < .025 \). On the other hand, for function words the difference is even larger for serial than for whole method of presentation, suggesting that for function words the immediate associations are more important than they are for the nonsense syllables. The mere separate (serial) presentation, even in sentence form, is sufficient to give rise to less recall for the nonsense syllables so that randomization does not lead to further decrement. In other words, the nonsense syllables may depend more upon long-range associations between words. Thus, Epstein's failure to find a difference in memory due to word order when using the serial method may be ascribed largely to the difficulty in recalling nonsense syllables under these conditions. The variable of word order is also involved in the triple interaction with grammatical structure and method of stimulus presentation. The interaction roughly corroborates Epstein's finding (1962) that the difference due to word order manifests itself when using the whole but not the serial method of presentation, as long as the grammatical structure consists of the D sentence only. His conclusion is not warranted, however, for the PNQ sentence which shows the difference due to grammatical structure to be at least as great for the serial as for the whole method of presentation. It might be noted here that the serial method as used in this experiment differs in two respects from Epstein's study. First, Ss were told that they would be presented words from a sentence in sentence or in random order; Epstein's Ss were not. Secondly, Ss responded in the same way for the whole and serial presentation methods, i.e., by writing out the entire sentence after seeing all stimuli one at a time, while Epstein's Ss followed the usual anticipatory serial learning instructions. Our findings suggest that Epstein's conclusion—"chains of immediate probabilistic associations within the structured sentences" cannot be used to explain the difference due to grammatical structure—may have been premature.

**Stage of Learning.** The Ss made a significantly larger number of errors on the first than on the second series of words learned, \( F(1, 288) = 45.59, p < .005 \). Furthermore it has already been indicated how the stage of learning interacts with word order and with grammatical structure.

Stage of learning is of course a frequency-of-occurrence variable, and the results suggest its importance for explaining differences in ability to recall various grammatical structures. Since the D structure most likely occurs quite frequently in verbal behavior (text as well as speech) to which the average S is exposed, one would expect more positive transfer to it than to
the PNQ structure which occurs less frequently. It is unlikely that this can be a warm-up effect since all Ss were exposed to a learning task before this task, namely the learning of a series of numbers.

Method of Stimulus Presentation. The whole method of presentation is superior to the serial method, $F(1, 288) = 13.51, p < .005$, but the interaction with stage of learning indicates (as mentioned above) that this difference occurs only for the first presentation. Other interactions of this variable have already been discussed above.

Word Type. Nonsense syllables are almost three times more difficult to memorize than function words, $F(1, 288) = 98.46, p < .005$. This difference is involved in a significant triple interaction with method of stimulus presentation and word order. Although the nonsense syllables give rise to higher error scores than do function words for all conditions, the difference is greater in sentences than in random arrangements. Although the function words are easier to recall than nonsense syllables because of their higher frequency of occurrence, they also show more dependence upon context than the nonsense syllables. The relatively greater dependence of function words upon context is in agreement with a recent experiment by Glanzer (1962), who has maintained that function words are incomplete units but can be made complete by embedding them among nonsense syllables which alone are also incomplete units. This experiment showed that function words can be made part of a complete unit by embedding them in a nonsense syllable sentence (for more extensive discussion of the problem of unit see Salzinger, 1962).

Discussion

Thus, in general we are led to the conclusion that Miller's notion (1962), that an S commits a complex sentence to memory by storing its kernel plus a footnote concerning the selection of the appropriate transformation, is not corroborated by the data of this experiment. Notions of frequency of occurrence (a basic variable in behavior theory) of the words or their combinations contained in the different structures appear to agree more with our findings. The importance of the frequency-of-occurrence variable continues to be demonstrated in current research. A recent example was given by Baddeley (1964) who showed that series of nonsense syllables can be learned more rapidly if the last letter of each syllable and the first of the subsequent one are "compatible" than when they are not. Compatibility is defined merely in terms of the number of letters required before an S will guess the next letter—a simple frequency-of-occurrence relationship.

It might also be pointed out that other experimenters have provided alternative formulations to explain the influence of grammatical variables in verbal behavior. Braine (1963a, 1963b) has suggested that people acquiring a language learn the location of units within sentences and associations between pairs of words. Jenkins and Palermo (1964) suggest sequence and class of words as important aspects of language learning. Finally, Salzinger (1959, 1965) and Staats (1961) pointed out the importance of the concept of response class in operant conditioning, i.e., what response members do in fact combine to form a response class, and in another paper Salzinger (1962) showed the importance of unit size for understanding the properties of speech. In summary, a number of investigators have suggested concepts derived from learning theory which could explain at least those properties of grammar currently being investigated. In view of the equivocal nature of the behavioral evi-
GRAMMAR AND RECALL

239
dence for a transformation grammar, it
would seem advisable at this point to in-
vestigate further the application of fre-
quency concepts to simple structures and
then determine what concepts, if any, must
be added to handle more complex struc-
tures. Such work has already begun in ex-
periments by Martin and Jones (1965) and
Martin, Davidson, and Williams (1965).

References

Baddeley, A. D. Language-habits, S-R compatibil-
ity, and verbal learning. Amer. J. Psychol.,
1964, 77, 463-468.

Bever, T. G., Fodor, J. A., and Weksel, W. On
the acquisition of syntax: a critique of "con-
textual generalization." Psychol. Rev., 1965,
72, 467-482.

Braine, M. D. S. On learning the grammatical
order of words. Psychol. Rev., 1963, 70, 323-
348. (a)

Braine, M. D. S. Learning language structures.
Paper presented at the American Psychol.
Assoc., Phila., Pa., 1963. (b)

Chomsky, N. Syntactic structures. 's Gravenhage:
Mouton and Co., 1957.

Epstein, W. The influence of syntactical structure
on learning. Amer. J. Psychol., 1961, 74,
80-85.

Epstein, W. A further study of the influence of
syntactical structure on learning. Amer. J.
Psychol., 1962, 75, 121-126.

Glanzer, M. Grammatical category: A rote learn-
ing and word association analysis. J. verb.

Glaze, J. A. The association value of nonsense
syllables. J. genet. Psychol., 1928, 35, 255-
267.

Jenkins, J. J., and Palermo, D. S. Mediation
processes and the acquisition of linguistic
1964, 29, 141-169.

Martin, J. G., Davidson, Judy R., and Williams,
Myrna L. Grammatical agreement and set
in learning at two age levels. J. exp. Psychol.,
1965, 70, 570-574.

Martin, J. G., and Jones, R. L. Size and struc-
ture of grammatical units in paired-associate
learning at two age levels. J. exp. Psychol.,
1965, 70, 407-411.

Mehler, J. Some effects of grammatical trans-
formations on the recall of English sentences.
J. verb. Learn. verb. Behav., 1963, 2, 346-
351.

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

Miller, G. A. Some psychological studies of
grammar. Amer. Psychol., 1962, 17, 748-762.

Miller, G. A., Galanter, E., and Pribram,
K. H. Plans and the structure of behavior.

Noble, C. E. Analysis of meaning. Psychol. Rev.,
1952, 59, 421-430.

Salzinger, K. Experimental manipulation of ver-
bal behavior: a review. J. gen. Psychol., 1959,
61, 65-94.

Salzinger, K. Some problems of response measure-
ment in verbal behavior: the response unit
and intraresponse relations. Paper presented
at the conference on Methods of Measurement
of Change in Human Behavior, Montreal,
Canada, 1962.

Salzinger, K. The problem of response class in
verbal behavior. Paper presented at the Con-
ference on Verbal Behavior, New York City,
1965.

Staats, A. W. Verbal habit-families, concepts,
and the operant conditioning of word classes.

Underwood, B. J. and Schulz, R. W. Meaning-
fulness and verbal learning. Philadelphia:

(Received June 21, 1965)