Effects of task on the perceptual organization of sentences*

PAUL R. AMMON, BERNARD OSTROWSKI†, and KEITH ALWARD
Department of Education, University of California, Berkeley, California 94720

Children and college students listened to sentences and responded to a question or a probe word after each sentence. To prevent Ss from using different perceptual strategies in performing the two tasks, question items and probe items were presented in mixed lists. In the probe task, the relationship between sentence structure and correct response latencies resembled the findings of an earlier study with block lists. The results from the question task were less similar to those obtained earlier, perhaps because, with mixed lists, Ss tended to organize all sentences in a manner especially suited to the probe task. A different type of organization might be used in the process of comprehending a sentence solely for the purpose of answering questions.

Psycholinguistic phenomena are sufficiently complex that a single experiment rarely produces results which are unequivocal. The results of an earlier study (Ammon, 1968) were interpreted as evidence that the process of perceiving grammatical relations in spoken sentences involves two separate levels of organization. The possibility of an alternate interpretation led to the present experiment, which may be introduced by a review of the earlier findings.

In the original study, Ss listened to five types of tape-recorded sentences, as shown below.

- (1) A noisy₁ squirrel₂ scolded the playful₃ children₄; he₅ climbed the shady tree.
- (2) A noisy₁ squirrel₂ scolded the playful₃ children₄; they₅ climbed the shady tree.
- (3) A noisy₁ squirrel₂ who₃ scolded the playful₄ children₅ climbed the shady tree.
- (4) A noisy₁ squirrel₂ scolded the playful₃ children₄ who₅ climbed the shady tree.
- (5) A noisy₁ squirrel₂ who₃ climbed the shady₄ tree₅ scolded the playful children.

A question such as "Who climbed the tree?" was presented immediately after each sentence type, and the time taken to produce a correct response was found to increase as a function of three variables: *separation* of grammatical subject and predicate (Types 1 and 3 vs Types 2 and 4); *form* of pronoun (Types 3 and 4 vs

Types 1 and 2); and *confusability* (Type 3 vs Type 5).

The same five sentence types were also presented with probe words. The time taken to produce the word which had followed the probe word in each sentence varied with the probe's position in the sentence (see Subscripts 1-5 in the above examples). These differences between probe positions could generally be predicted from a tree diagram of surface phrase structure (see Ammon, 1968). With phrase structure held constant, the variables called *separation* and *confusability* were found to have no effect on the profile of response latencies across probe positions. These results led to the suggestion that probe responses reflect a separate level of organization in the process of comprehension, a level which corresponds to surface phrase structure and is relatively independent of deeper grammatical analysis.

An important feature of the original study was the fact that the Ss performed the question and probe tasks on block lists, responding to questions in one part of the experiment and to probe words in another. It seems quite likely that the use of block lists might induce a differential listening strategy, depending on the task to be performed. The possibility of a task-dependent strategy makes it difficult to claim that responses to questions and probes reflect separate levels of organization in the same process of comprehension. An obvious remedy to the problem is to mix question items and probe items in the same list of sentences. With a mixed list, the Ss would not know which task they had to perform until after they had heard each stimulus sentence. Under these conditions, it is reasonable to assume that question items and probe items would be processed in the same way, at least up until the presentation of the question

or probe word following the sentence. The purpose of the present experiment was, therefore, to see whether or not the use of mixed lists would produce different results from those obtained with block lists in the original study.

METHOD

Subjects

The Ss were 20 fourth graders and 20 college students, with each age group divided equally as to sex. The fourth-grade sample was drawn from an elementary school serving a predominantly middle-class community in the San Francisco Bay Area. The college Ss were enrolled in psychology and education courses at the University of California, Berkeley.

Stimulus Materials and Procedure

The stimulus items and the testing procedure were identical to those used by Ammon (1968), except that question items and probe items were mixed in a random order. Each S was tested in two sessions at least one day apart. In the first session, S heard 100 sentences, half with questions and half with probes. The same sentences were presented again in the second session, but the previously probed items were questions, and vice versa.

As in the original study, the question items of most interest were those involving a question like "Who climbed the tree?" (Q 1). But in order to assess the effects of *confusability*, it was also necessary to consider questions like "Who scolded someone?" (Q 2)—see Examples 3 and 5 above.

Treatment of Data

The data were prepared for analysis in much the same way as before (Ammon, 1968), except that response latency was measured (in millimeters on paper charts) from the onset rather than the end of the question or probe word. The raw latencies were converted to their reciprocals, multiplied by 1,000, and rounded to the nearest whole number. Then the mean of these scores was computed for the correct responses produced by each S under each condition of the experiment. These cell means for individual Ss will be referred to as transformed latency (TL) scores.

RESULTS

Effects of Separation and Form on Question Responses

Table 1 shows the mean TL score for Sentence Types 1-4 as a function of separation, form, age, and sex. An analysis of variance was performed on these data with a 2 by 2 by 2 by 2 factorial design. The only significant effect associated with the stimulus variables was the interaction of

*This research was supported by Small Grant MH 14685-01 from the National Institute of Mental Health and by funds from the Committee on Research, University of California, Berkeley. The present report was prepared at the University of California's Institute of Human Learning, which is supported by grants from the National Institutes of Health. The authors thank Kay Byers for her assistance in analyzing the data.

†Now at the University of Oregon.

Table 1
Mean TL Score for Correct Responses to Question 1 as
a Function of Separation, Form, Age, and Sex

Sex	Form	Fourth Grade		College	
		Separation	Separation	Separation	Separation
Female	Personal	16.8	16.5	21.3	21.4
	Relative	15.6	16.8	20.0	20.9
Male	Personal	18.0	15.9	17.7	17.7
	Relative	17.2	15.9	18.2	17.7

Table 2
Mean TL Score for Correct Responses to Questions 1 and 2
as a Function of Confusability, Age, and Sex

Sex	Question	Fourth Grade		College	
		Confusability	Confusability	Confusability	Confusability
Female	Q 1	15.6	17.6	20.0	20.9
	Q 2	15.9	15.1	21.0	21.3
Male	Q 1	17.2	16.3	18.2	19.6
	Q 2	15.5	15.7	19.2	19.2

separation with sex, $F(1,36) = 5.10$, $p < .05$. Females tended to replicate the original finding of a difference in favor of unseparated sentences, whereas males showed a sizable difference in the opposite direction. As in the earlier study, higher TL scores occurred with the personal pronoun form vs the relative pronoun form, but the present difference fell short of significance, $F(1,36) = 1.11$. The only other significant result of this analysis was the main effect of age, $F(1,36) = 7.42$, $p < .01$.

Effects of Confusability on Question Responses

Type 3 sentences were designated as more confusable than Type 5 sentences. But, with respect to Q 1, confusability was confounded with separation. For this reason, it was necessary to include Q 2 in the analyses of the confusability effect. Thus, if there were a main effect of confusability (Type 3 vs Type 5), the interaction of confusability with question could be examined to determine whether or not the confusability effect might have occurred only with Q 1, in which case it could be interpreted as an artifact of separation. The obtained mean TL scores appear in Table 2 as a function of confusability, question, age, and sex. Another analysis of variance with a 2 by 2 by 2 factorial design showed that only the usual effect of age was significant, $F(1,36) = 15.03$, $p < .001$. The effect of confusability was in the expected direction but was nonsignificant, $F(1,36) < 1$, as was the interaction of confusability with question, $F(1,36) = 1.61$.

Effects of Separation, Confusability, and Probe Position on Probe Responses

Of the 20 fourth-grade Ss, 15 failed to produce a correct response in one or more of the conditions of the experiment (usually more than one), precluding an analysis of their data with respect to the latencies of correct responses. Thus, only the data for college Ss are presented in Table 3.

Sentence Types 1 and 2 differed in terms of separation. The data from these two types were subjected to a

2 by 5 by 2 analysis of variance, in which the factors were separation, probe position, and sex. The analysis showed no significant main effect of separation, but there was a significant interaction of separation with sex, $F(1,18) = 8.29$, $p < .01$. The female Ss had a clear difference in favor of unseparated sentences—as in the original study—while the male Ss had a slight difference in the opposite direction. The absence of a significant Separation by Probe interaction was also similar to the earlier study.

Because Types 1 and 2 shared a common surface structure, and because the interaction of sentence type with probe position was not significant, it was possible to examine the differences between probe positions in relation to phrase structure. The main effect of probe position was significant, $F(4,72) = 78.73$, $p < .001$. Pairwise Scheffé contrasts with a critical value of 4.8 showed that the TL scores in Positions 1, 3, and 5 were significantly higher than those in Positions 2 and 4, at the .05 level. There was also significant interaction of probe position with sex, $F(4,72) = 11.06$, $p < .001$. The difference in favor of female Ss got progressively larger toward the end of the sentence. Nevertheless, the profile of TL scores across probe positions seems about the same for both sexes.

A second analysis was based on the data from Types 3 and 5, which also shared a common phrase structure. Confusability, probe position, and sex were the independent variables in

another 2 by 5 by 2 factorial design. There were no significant effects associated with confusability. There was, however, a main effect of sex, with females attaining higher TL scores than males, $F(1,18) = 8.45$, $p < .01$. Unlike the preceding analysis, sex did not interact significantly with probe position. As for probe position per se, the main effect was again significant, $F(4,72) = 56.02$, $p < .001$. The critical value for pairwise contrasts was 6.0, so that the means for Positions 1 and 3 were significantly higher than those at Positions 2, 4, and 5.

Finally, the data for Sentence Type 4 were analyzed with respect to the effects of probe position and sex in a 5 by 2 factorial design. The females again outscored the males— $F(1,18) = 6.27$, $p < .05$ —and the Sex by Probe interaction was not significant. There was a significant effect of probe position, $F(4,72) = 19.06$, $p < .001$. Pairwise contrasts with a critical value of 5.3 indicated that the means for Probes 1 and 3 were significantly higher than those for Probes 2 and 4, and that Probe 5 had a higher mean than Probe 2.

As in the original study, all of the present significant contrasts between probe positions were predicted from tree diagrams of surface phrase structure, but some of the predicted differences were not significant. The pattern of significant contrasts differed slightly in the present experiment, due to the fact that Probe 5 had a lower mean relative to

Table 3
Mean TL Score for Correct Responses to Probe Words as a Function of Sentence
Type, Probe Position, and Sex (College Ss Only)

Sex	Type	Probe Position				
		1	2	3	4	5
Female	1 (Separation)	32.0	22.5	34.1	20.3	38.0
	2 (Separation)	32.8	24.2	35.5	24.6	39.5
	3 (Confusability)	36.2	21.1	36.5	20.8	24.4
	4	36.8	26.1	35.0	27.4	31.3
	5 (Confusability)	33.9	22.3	34.1	20.6	26.1
Male	1 (Separation)	33.5	20.9	31.6	17.0	31.9
	2 (Separation)	31.1	21.1	29.4	19.4	31.7
	3 (Confusability)	31.6	16.8	28.9	15.5	16.5
	4	29.8	17.6	32.2	22.1	26.3
	5 (Confusability)	30.2	16.9	29.4	17.0	20.0

Fig. 1. Mean TL scores for college Ss in probe task, comparing block lists (Ammon, 1968) and mixed lists (present experiment).

the other probes in Types 3 and 5, and due to some differences in Type 4 which only approached significance in the present data. The general similarity between the present probe results and those obtained in the original study may be seen graphically in Fig. 1. The remarkable thing about these graphs is the extent to which the two profiles in each plot are parallel.

DISCUSSION

This experiment with the probe and question tasks mixed in the same list of sentences was conducted for the sake of comparison with results obtained in an earlier study using block lists (Ammon, 1968). There are, however, a few limitations on the comparability of the two sets of data. The two studies were conducted in different parts of the country. Although the samples were roughly equivalent with respect to age and education, the populations from which they were drawn may differ in terms of variables which influence performance of the experimental tasks. One obvious difference in the samples was the inclusion of males in the present experiment, whereas the original study had only female Ss. It still is possible to compare the data for females in the two experiments, but the number of females in the present experiment was only half that of the original study, resulting in a loss of statistical power. On the other hand, the use of male Ss turned up some unexpected sex differences.

To summarize the results of the question task, the original finding of a

separation effect received weak support only for females. The form effect found earlier was not replicated, perhaps because the use of mixed lists forced the Ss to attend more closely to surface details of the sentences. A similar phenomenon may have weakened the effect of confusability, which was also not significant in the present study.

With regard to the probe task, the old and new data for college Ss seem remarkably similar. As in the original study, there was evidence of a main effect of separation for female Ss, but no interaction of separation with probe position, and no significant effect of confusability. As for the profiles of latency across probe positions, Fig. 1 illustrates the extent to which the original findings were replicated, despite the use of mixed lists. The differences in level between the two sets of data in Fig. 1 can be attributed, in part, to the measurement of raw latencies from the onset rather than the end of the probe word in the present experiment. It seems likely, too, that the use of mixed lists resulted in some extra decision time during which Ss identified the task to be performed on each sentence. This extra step may have made the probe task too difficult for most of the younger Ss in the present experiment.

The absence of significant interactions between age and stimulus variables (where these interactions could be tested) was consistent with the earlier findings. However, in both

the question and the probe results, separation was found to interact significantly with sex, such that the expected effect of separation on response latency was, if anything, reversed in the case of males. There was also a Sex by Probe interaction in which female Ss responded more quickly than males to probe words taken from nearer the end of the sentence. These three interactions might reflect the same underlying mechanism—namely that males tend to focus more on the beginning of the sentence than do females. Such a primacy effect would reverse the effect of separation on males, because the response word for a separated predication would be at the beginning of a sentence. The effects of stimulus variables should be tested separately for males and females in future experiments.

In conclusion, the present study offers little reason to doubt that the probe task reflects a type of organization in sentence perception which corresponds to surface structure and is more or less independent of deeper grammatical analysis. It still is not clear, however, that the same degree of independence occurs in the normal process of comprehension in view of the present results from the question task. The effects of sentence structure on responses to questions about deeper grammatical relations generally were not replicated with mixed lists. Perhaps the Ss—both male and female—tended to process all sentences in the mixed lists as they would for the probe task alone. This strategy would make sense, in that the deeper grammatical relations needed to answer questions could be derived from a verbatim organization of the sentence, but not vice versa. The verbatim type of organization would probably play less of a role with a block list of question items, or in everyday comprehension. Thus, when the question and probe tasks are performed on block lists, they may very well be revealing alternate perceptual strategies rather than separate levels of organization in the same process of sentence perception.

REFERENCE

- AMMON, P. R. The perception of grammatical relations in sentences: A methodological exploration. *Journal of Verbal Learning & Verbal Behavior*, 1968, 7, 869-875.

(Accepted for publication April 3, 1971.)