COGNITIVE PSYCHOLOGY 14, 178–210 (1982)

Making and Correcting Errors during Sentence Comprehension: Eye Movements in the Analysis of Structurally Ambiguous Sentences

LYN FRAZIER AND KEITH RAYNER
University of Massachusetts

Eye movements were recorded as subjects read sentences containing temporary structural ambiguities. In accord with the garden-path theory of sentence comprehension, shorter reading times were found for sentences conforming to certain independently motivated parsing strategies (late closure and minimal attachment) than for comparable sentences which violate these strategies. Further, longer fixation durations were associated with the very first fixation in the region of the sentence which disambiguated the sentence, suggesting that the human sentence-parsing mechanism operates in a rather systematic fashion, immediately computing the structural consequences of fixated material for the analysis of preceding material. The pattern of regressive eye movements did not conform to the view that the parsing mechanism automatically returns to the beginning of the sentence to revise an incorrect analysis of linguistic material nor did it support the view that the parsing mechanism systematically backtracks through the sentence until the source of the erroneous analysis is located. Rather, the pattern of regressions indicated that the parsing mechanism typically engages in selective reanalysis, exploiting whatever information it has available about the type of error it has committed to guide its reanalysis attempts. Finally, it is emphasized that an understanding of the parser's revision procedures is essential to an explanation of why certain linguistic structures cannot be successfully parsed by humans.

In recent years a variety of approaches have been taken to the question of how people understand the sentences of a natural language. One approach to this question is what might be called the garden-path theory of sentence comprehension. According to this view, the human sentence-parsing mechanism (hereafter the parser) copes with the temporary ambiguities of natural language by initially pursuing just a single analysis of a sentence. Of course, if there is more than one permissible analysis of some portion of the sentence, then there is no guarantee that the analysis which was chosen will turn out to be correct. And therefore the parser will often be led down the garden-path.

This research was supported by grant BNS79-17600 from the National Science Foundation. We would like to thank Maria Slowiaczek and Jim Bertera for their assistance in collecting the data and Arnold Well, A. W. Inhoff, Charles Clifton, Jr., and two anonymous reviewers for their comments on a previous draft of the paper. Requests for reprints should be addressed to Lyn Frazier, Department of Linguistics, South College, University of Massachusetts, Amherst, Massachusetts, 01003.

0010-0285/82/020178-33$05.00/0

In this study we explore the garden-path theory of sentence comprehension using data from people's eye movements during the reading of structurally ambiguous sentences. Specifically, we argue that a small number of identifiable patterns characterize people's recovery from an initial misanalysis of a sentence.

The garden-path approach to sentence comprehension may be contrasted with the view that all possible analyses of linguistic material are pursued simultaneously (see discussion of the Parallel Processing Hypothesis in Fodor, Bever, & Garrett, 1974) and with the Minimal Commitment Hypothesis which claims that the parser typically delays making a decision about temporarily ambiguous material until it has received sufficient information to disambiguate the sentence (for one version of this hypothesis see Marcus, 1980). These views suggest that the only garden paths or systematic errors in sentence analysis are those of which people are consciously aware (as in the sentence *The horse raced past the barn fell*, where people report having misinterpreted the string *The horse raced past the barn* as if it were a simple active sentence). Since such conscious errors of analysis are relatively rare, the garden-path phenomenon is assumed to be a marginal phenomenon, certainly not characteristic of the parser's fundamental response to the pervasive temporary ambiguities of natural languages.

The garden-path theory, however, suggests that garden-path phenomena account for many common, minor disruptions of sentence comprehension. Presumably, the parser obeys systematic decision preferences in analyzing ambiguous material. If the parser's first analysis of an early portion of a sentence happens to be compatible with later disambiguating material, then the sentence should be relatively easy to process. However, if the disambiguating material should prove to be incompatible with the parser's initial analysis of preceding material, then the sentence should be relatively difficult to process since the parser will have to revise its original analysis of the sentence.

For example, if the parser prefers to structure incoming lexical material together with material that has already been received (rather than structure it with subsequent material, i.e., as a sister to the following phrase), then the parser will analyze the temporarily ambiguous phrase *a mile* in (1) as the direct object of the verb *jogs*, not as as the subject of the following clause. Hence, the complexity of processing a sentence beginning with (1) will depend on whether the remaining portion of the sentence confirms this initial decision (as in (1a)), or refutes it (as in (1b)).

(1) Since Jay always jogs a mile . . .
 (a) Since Jay always jogs a mile this seems like a short distance to him.
 (b) Since Jay always jogs a mile seems like a short distance to him.

This sort of asymmetry in the processing complexity of temporarily ambiguous sentences provides the strongest evidence for the garden-path theory of sentence comprehension. Both the parallel processing hypothesis and the minimal commitment hypothesis can account for an increase in processing complexity associated with the presence of a temporary ambiguity. The parallel processing hypothesis may simply attribute the increase in processing complexity to the additional computations required to develop more than one analysis of the material, while the minimal commitment hypothesis may attribute the increase in perceptual complexity to the need to hold information in memory while delaying certain decisions. What these hypotheses cannot account for is an asymmetry in the complexity of the two analyses of temporarily ambiguous material.

Frazier (1978) reviews the evidence for the garden-path theory of sentence comprehension and argues that many of the decision principles proposed in the psycholinguistic literature can be subsumed under two very general parsing strategies, *late closure* and *minimal attachment*.[1]

Late closure: When possible, attach incoming lexical items into the clause or phrase currently being processed (i.e., the lowest possible nonterminal node dominating the last item analyzed).

Minimal attachment: Attach incoming material into the phrase-marker being constructed using the fewest nodes consistent with the well-formedness rules of the language.

The late closure strategy was illustrated above in the discussion of sentence (1). Late closure predicts that the temporarily ambiguous noun phrase *a mile* will initially be analyzed as the direct object of the verb *jogs* since this permits it to be analyzed as a constituent of the (verb-)phrase currently being parsed. Hence late closure predicts that sentence (1a) should be easier to process than sentence (1b) where this analysis will turn out to be incompatible with subsequent context.

The minimal attachment strategy will apply in the processing of a sentence fragment such as that shown in (2). It predicts that the temporarily ambiguous noun phrase *the mayor's position* will be interpreted as the simple direct object of the verb *argue* as indicated in (2a), rather than as the subject of a sentential complement as in (2b), since the former analysis requires the postulation of fewer nodes (see Figs. 1–2).

(2) The city council argued the mayor's position . . .
  (a) The city council argued the mayor's position forcefully.
  (b) The city council argued the mayor's position was incorrect.

[1] See Frazier (1978) for a discussion of these strategies and their relation to Kimball's (1973) strategies. The explanation for why different people should exhibit these same parsing preferences and the specific model of sentence comprehension in which these strategies were developed are discussed in Frazier and Fodor (1978) and Fodor and Frazier (1980).

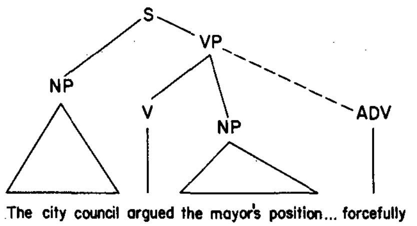

FIG. 1

Frazier (1978) tested the predictions of these strategies in a number of different constructions using serial visual presentation and a grammaticality judgment task. The outcome of those experiments clearly confirmed the predictions of the strategies and thus supported the garden-path theory of sentence comprehension.

If people do frequently commit themselves to analyses which they later must revise, then it is important to study the processes they employ to reanalyze sentences. These revision procedures bear much of the burden of predicting the overall processing complexity of sentences. Further, characterizing the class of "unparsable" linguistic structures will presuppose an understanding of these procedures since a sentence is incomprehensible only if it defies both the parser's first pass analysis attempts and its attempts at reanalysis. Hence, if psycholinguistic theory is to be successful in determining the boundary constraints which the human sentence-parsing apparatus imposes on the grammars of natural languages (i.e., on the possible or at least usable sentences of the language), it will be necessary to identify the characteristics which render a misanalyzed construction impervious to the parser's normal (unconscious) reanalysis procedures.

Though the correction routines used to revise an initial misanalysis of a sentence have not been a major focus of psycholinguistic research, there are a few suggestions on this topic which can be culled from the psycholinguistic literature. One obvious hypothesis, which seems to have been implicit in some of the earlier discussions of garden-path sentences, is that the parser returns to the very beginning of a sentence and processes

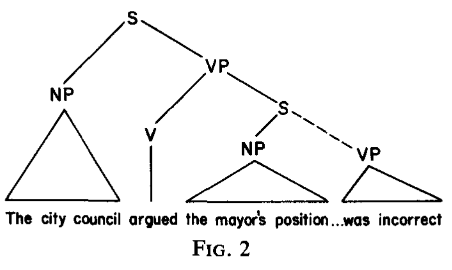

FIG. 2

the entire sentence over again (looking for alternative decisions at choice points) whenever it detects an error in its first analysis. This *forward reanalysis hypothesis* seems to correspond to one's intuitions concerning the conscious reanalysis of sentences (but, as we shall suggest below, this may simply be because in cases of conscious reanalysis the parser's normal correction routines have been unable to correctly diagnose the source of the error and thus the parser may be forced to abandon its normal revision procedures).

A second alternative, the *backward reanalysis hypothesis*, has been suggested within the ATN framework (cf. Kaplan, 1972). As suggested by its name, this hypothesis predicts that the parser proceeds backward from the point of breakdown in its initial analysis, systematically retracing its initial decisions and trying out alternatives.

A third possibility is what we shall call the *selective reanalysis hypothesis*. This hypothesis was suggested in one form by Winograd (1972), who proposed that the language comprehension system uses whatever information is at its disposal to help in any necessary restructuring of linguistic material. Kimball's (1973) statement of the fixed structure principle (which claims that there is a cost associated with pulling a shunted phrase out of memory to restructure it) also seems to presuppose the existence of some sort of selective reanalysis since it implies that reanalysis can be sufficiently localized to affect just a specific phrase. Carpenter and Daneman (1981) suggest a similar hypothesis with respect to the processing of lexical ambiguity.

Though the selective reanalysis hypothesis needs considerable refinement before it will make precise claims about the parser's correction procedures, the essence of the hypothesis is the claim that the parser does not stupidly and automatically proceed through the sentence in one direction or the other regardless of the type of error involved; rather, the parser will use whatever information indicates that its initial analysis is inappropriate to attempt to diagnose the source of its error. If successful, this would permit it to selectively focus on just that portion of the analysis which was responsible for the particular problem it encountered with its first analysis. The selective reanalysis hypothesis thus predicts that the correction procedures which are involved in the reanalysis of a sentence may be very specific and will depend on the exact nature of the evidence which the parser has available to it.

One virtue of the selective reanalysis hypothesis is that it would explain why some sentences appear to be much easier to reanalyze than others. For example, in a sentence such as *The linguists knew the solution to the problem would not be easy*, minimal attachment predicts that the parser will initially construct the phrase marker shown in Fig. 3. When the phrase *would not* . . . is received, there will be no legitimate attachment

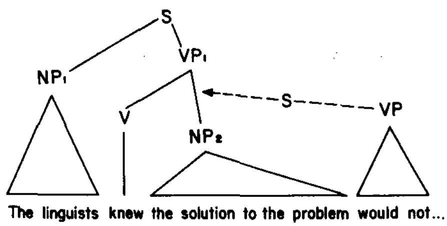

FIG. 3

site available for this phrase. The parser could, however, determine that this verb phrase is missing its subject and presumably this information, together with the fact that the preceding noun phrase would be a legitimate subject for this verb phrase, could be used to arrive at the correct sentential complement analysis of the sentence (i.e., to insert an S node between NP_2 and VP_1). In the sentence *The horse raced past the barn fell*, minimal attachment predicts the parser will have constructed the analysis indicated in Fig. 4 before it receives the verb *fell*. Again it will receive a verb phrase (*fell*) for which no legitimate attachment will be available. In contrast to Fig. 3, however, there will not be any already identified noun phrase in Fig. 4 which could legitimately serve as the subject of this verb phrase. Thus, by hypothesis, the parser will incorrectly focus its attention on the problem of finding an attachment site for the verb *fell* in Fig. 4. Since the preceding material will have been assigned a perfectly sensible analysis, the parser will not have any readily identifiable clues that the true source of its error occurred much earlier in the sentence in its premature closure of the subject noun phrase. In short, we are suggesting that a sentence such as that in Fig. 4 is extremely problematic for the human sentence-parsing mechanism because it garden-paths both the parser's first pass analysis and also garden-paths its attempts at normal reanalysis.

To test the predictions of these various classes of reanalysis hypotheses, we measured subjects' eye movements and fixation durations as they read sentences such as those in Figs. 1–3. We considered the recording of eye movements to be an ideal technique for studying reanalysis

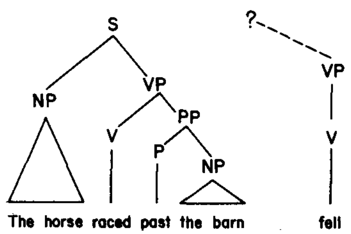

FIG. 4

routines since it provides a nonobtrusive on-line measure of processing complexity which could indicate local variations in processing time associated with the particular region of the sentence we are interested in.

METHOD

Subjects. Sixteen undergraduate students at the University of Massachusetts were paid to participate in the experiment. All of the subjects were naive with respect to the purpose of the experiment and they all had normal, uncorrected vision.

Materials. Sixteen closure and 16 attachment sentences were constructed, with four versions of each. The closure sentences consisted of an initial clause followed by one or more subsequent clauses. In half the closure sentences, the initial clause was a subordinate clause followed by a single main clause. The remaining closure sentences had three conjoined clauses. As indicated in (5) below, the initial clause of a closure sentence always contained a temporarily ambiguous postverbal noun phrase (underlined in (5)). According to the late closure strategy, this noun phrase would initially be analyzed as the direct object of the preceding verb. In the late closure (LC) version of a sentence, this decision would in fact lead to the correct analysis of the sentence. In the early closure (EC) version of the sentence, this decision would turn out to be incorrect, since the temporarily ambiguous noun phrase is in fact the subject of the following clause.

(5) LC-Long: Since Jay always jogs *a mile and a half* this seems like a short distance to him.
EC-Long: Since Jay always jogs *a mile and a half* seems like a very short distance to him.
LC-Short: Since Jay always jogs *a mile* this seems like a short distance to him.
EC-Short: Since Jay always jogs *a mile* seems like a very short distance to him.

The length of the temporarily ambiguous noun phrase was also varied: short ambiguous phrases consisted of just one or two words; long ambiguous phrases contained three or more words (usually four or five). Thus, for each of the 16 closure sentences, there were two early closure versions (containing either a short or long ambiguous phrase) and two late closure versions (again, with either a short or long ambiguous phrase).

Similarly, there were four versions of each of the 16 attachment sentences, as indicated in (6) below. The attachment sentences also contained a temporarily ambiguous postverbal noun phrase. According to the minimal attachment strategy, this noun phrase would initially be interpreted as the simple direct object of the preceding verb.[2] In the minimal attachment (MA) version of a sentence, this noun phrase is in fact the direct object of the preceding verb; whereas, in the nonminimal attachment (NM) version of the sentence, the postverbal noun phrase is actually the subject of an embedded sentential complement to the verb.

(6) MA-Long: The lawyers think his second wife will claim *the entire family inheritance.*
NM-Long: The second wife will claim *the entire family inheritance* belongs to her.
MA-Short: The lawyers think his second wife will claim *the inheritance.*
NM-Short: The second wife will claim *the inheritance* belongs to her.

The length of the ambiguous noun phrase was varied in the attachment sentences (as in the closure sentences) yielding a long and short version of both the minimal attachment and nonminimal attachment sentences. To ensure that the different versions of a sentence had

[2] Although at a superficial level the late closure and minimal attachment strategies appear to be very similar, in reality they cannot be reduced to a single strategy. See Frazier (1978) for arguments.

the same number of clauses, it was necessary to embed the clause containing the ambiguous phrase in a higher clause in the minimal attachment sentences. This manipulation necessitated the use of different lexical items in the two versions of these sentences. In other respects, the lexical content of the sentences was matched as far as possible.

The four versions of each closure and attachment sentence were divided into four sets of sentences with each set containing one version. These four sets contained an equal number of each version; two different versions of a single sentence never appeared within the same set. This counterbalancing procedure ensured that each subject read four instances of each of the four versions of the closure and attachment sentences. In addition to the 16 closure and 16 attachment sentences, 72 filler sentences were constructed. These sentences had a wide variety of different structures. Ten of the filler sentences served as warmup sentences and the remaining 62 filler sentences, 16 closure, and 16 attachment sentences were randomized within the four different sets of sentences. The appendix lists the closure and attachment sentences.

We should emphasize that none of the sentences used in the experiment contained commas. The closure sentences deserve some comment since a comma would often be used to disambiguate these sentences in normal prose. Our reason for excluding sentence internal punctuation in the sentences stems from our desire to separate out and identify the various lexical, syntactic, and semantic routines used during normal sentence comprehension by depriving the parser of certain types of information before attempting to study the interaction of the routines.

This research strategy obviously has its pitfalls and would not seem well motivated in cases where removing certain information from a sentence actually forced the parser to cope with the sentence in a way that is totally unrelated to its usual mode of operation. But, in this case, we would not really expect systematic behavior across different subjects, since each would be coping with an unusual situation as best he could. Further, in the present case, punctuation which might disambiguate the closure sentences is in many instances optional. Moreover, the late closure strategy has already been shown to operate in numerous other examples beyond those where punctuation could help to disambiguate the sentence. For example, late closure predicts that the adverb *yesterday* will be attached into the lower clause in the sentence *John said Mary left yesterday*, where it will modify the verb left, rather than being attached higher, where it would modify the verb *said*. Of course, the issue of punctuation does not even arise in these cases, or in the case of the attachment sentences, where the presence of sentence-internal punctuation would be completely inappropriate.

Apparatus. The sentences were presented on a Hewlett–Packard cathode ray tube (CRT) which was interfaced with a Hewlett–Packard 2100A computer. The sentences, which varied in length from 8 to 28 words, were presented on 1–4 lines of double-spaced lower-case type font.

Eye movements were recorded via a Stanford Research Institute Dual–Purkinje Eyetracker (Cornsweet & Crane, 1973) which has a resolution of 10 min of arc. The horizontal and vertical signals from the eyetracker were sampled every millisecond by the computer, and eye position was determined by comparing the signals each 4 msec with the prior 4 msec. A complete record of eye location, fixation duration, saccade length, and fixation sequence was stored on the computer disk for each sentence.

Procedure. When the subject arrived for the experiment, a bite bar was prepared and served to eliminate head movements during the experiment. Then the eye-movement-recording system was calibrated for each subject. The initial subject calibration was accomplished in 5 to 15 min. After the initial calibration, the subjects were instructed that sentences were to be displayed on the CRT and that they were to read them for understanding. They were also informed that they would periodically be asked comprehension questions about the sentences. They were told that when they understood the sentence they were to push a button located near their right hand and that the button push would terminate the sentence. Prior to each sentence being read, a calibration check was carried out to make sure the eye movement recording system was appropriately aligned.

Each subject read 10 warm-up sentences followed by the set of 94 appropriate experimental and filler sentences. The experimenter asked the subject to release the bite bar on the average of every 10 sentences to answer a question about some arbitrary aspect of the last sentence read, although in practice it was quite random as to when questions were asked so that subjects could not anticipate which sentences they would be asked to respond to. Subjects were able to answer the questions without difficulty and were virtually always correct. The experiment typically required 2 hr of running time and subjects were allowed to take breaks whenever they desired. The subject's eye was 46 cm from the CRT and 3 characters equaled 1° of visual angle. Eye movements were monitored from the right eye and viewing was binocular. The room was dark, except for a very dim indirect light source that enabled the experimenter to read the comprehension questions to the subject. The CRT was adjusted to a comfortable brightness for each subject. Track losses due to pupil reactions and blinks occurred on 10% of the trials (range across subjects was 0–22%). Track losses resulted in inaccurate information about eye position and such trials were eliminated from the data analysis.

RESULTS

Prior to discussing the results of the data analyses which were performed, we will discuss problems associated with scoring the data and data analysis in general and then specifically describe predictions regarding reading time that can be generated from the different theoretical positions outlined in the introduction. Then we will discuss the results of the analysis of the closure sentences, the analyses of the attachment sentences, and an analysis concerning the pattern of regressions.

Data Analysis

For each sentence that a subject read, a vast amount of data was obtained consisting of the sequence and duration of each eye fixation. While the eye movement data are valuable because they allow one to examine local variations in processing difficulty which is not possible with global measures such as total reading time, they also present problems in that it may not always be apparent as to how to analyze such a vast amount of data. Accordingly, we have scored the data in a number of different ways and we will report the results of these analyses moving from the most global to the most local.

The most global analysis that we carried out was for reading time. However, inasmuch as the sentences differed widely in terms of the number of words making up the sentence, we have divided the total reading time per sentence by the number of character spaces (including spaces between words) in the sentence yielding *total reading time per letter.* In addition, so that we could more precisely examine specific predictions of the different hypotheses under consideration, we have also computed reading time per letter for different regions of the sentence. Furthermore, since reading time in different regions of the sentence may vary as a function of whether or not the ambiguity has been detected or resolved, we computed these local reading times for the *first pass* through the sentence versus the *second pass*. The first pass was defined as all fixations on a region of the sentence that had not already been read. The second pass, by definition, included all subsequent fixations in any region to the left of the rightmost or most advanced fixation. Occasionally, (about 4% of the time) subjects reread a sentence three or four times. All such rereadings were included as second pass readings. The crucial regions of the sentence for which we computed reading time were the regions *prior to the ambiguous region*, the *ambiguous region*, and the *disambiguating region* (the first two words immediately following the ambiguous phrase). The most local analysis which we performed consisted of the average fixation duration of the last three fixations prior to the disambiguating region and the first three fixations in that region. The final type of analysis that we will report is related to the *pattern of eye movements*. In particular, the frequency of regressive eye movements from one particular region of the sentence to another region was tabulated. For the regression-type analysis, regressions initiated from a region of the sentence prior to the disambiguating region were excluded, as were all within word regressions. Eight percent of the saccades prior to the disambiguating region were regressions. For similar procedures for analyzing eye movement data in experiments dealing with language processing, see McConkie, Hogaboam, Wolverton, Zola and Lucas (1979) and Just and Carpenter (1980).

Predictions

The hypotheses discussed in the introduction make a series of predictions concerning the processing of the experimental sentences. First, the *garden-path* theory (together with late closure and minimal attachment) predicts that, overall, reading times should be longer for the early closure and nonminimal attachment sentences than for late closure and minimal attachment sentences. Further, the increased processing times associated with the early closure and nonminimal attachment sentences should occur at (or slightly after) the point when subjects read the disambiguating material. Finally, if subjects detect the ambiguity in the experimental sentences, then we would expect reading time in the ambiguous region of the sentence to be longer than in the unambiguous region (i.e., before the ambiguity) even on the initial pass through the ambiguous region of the sentence. (A stronger test of this prediction would involve comparing reading times in the ambiguous region with the reading time for unambiguous material in the same position of the sentence.)

In contrast to the predictions of the garden-path theory, the parallel processing hypothesis predicts no difference in the overall processing times for the early closure and late closure version or the minimal attachment and nonminimal attachment version of a sentence. Instead, it predicts increased reading times in the ambiguous region of all sentences, due to the computations associated with developing more than one analysis of the material in this region of the sentence. Similarly, the minimal commitment hypothesis predicts no overall differences in reading times for the different closure or attachment versions of the sentences. Rather, it would most naturally predict longer reading time in the disambiguating region of all of the experimental sentences, since at this point in the processing of the sentence the parser will receive sufficient information to permit it to structure the preceding material (i.e., the ambiguous phrase).

Clearly we are considering the predictions of the simplest, most extreme versions of the parallel processing and minimal commitment hypotheses. The predictions of the parallel processing hypothesis might be modified by assuming that rejecting one analysis of the ambiguous phrase takes longer than rejecting the alternative. The predictions of the minimal commitment hypothesis might be modified by assuming a difference in the overall cost associated with the construction of the different analyses of the sentences. Though auxillary assumptions of this type might be devised to account for some set of data, these assumptions do not follow from the above hypotheses and would not make any predictions about the direction of any asymmetry in the perceptual complexity of our experimental sentences until they are cast in the form of some concrete proposal. Even then, the above predictions concerning the major locus of the perceptual difficulty associated with the ambiguous phrase would seem to remain intact.

The forward reanalysis hypothesis predicts that any regressive eye movements that occur should return to the very beginning of the sentence and then proceed forward through the sentence. The backward reanalysis hypothesis predicts that regressive eye movements should proceed backward through the sentence, regardless of where in the sentence the regressive eye movements were initiated. The selective reanalysis hypothesis, on the other hand, predicts that regressive eye movements will return directly to the ambiguous phrase, provided that the regression begins from the disambiguating region (i.e., the region containing the information which would permit the parser to locate the source of its error).

One reason for including the long ambiguous phrases in the experiment was to explore the parser's decision lag. If the parser's decisions about how to structure words with surrounding context are not made as soon as the words are fixated, then we might expect disambiguating information from subsequent words to influence the parser's initial analysis of earlier material. A decision lag of one or two words might be sufficient to prevent a garden-path effect in the early closure and nonminimal sentences with short ambiguous phrases, but not in the sentences with the long ambiguous phrases. Though a decision lag of one or two words could be distinguished from the minimal commitment hypothesis, a decison lag which was sufficiently long to incorporate all of the words of the long ambiguous phrases would be indistinguishable in our experiment from the predictions of the minimal commitment hypothesis discussed above. Should our results suggest a decision lag of this magnitude, we are quite prepared to take this as disconfirmation of the garden-path theory of sentence comprehension, rather than retreating to the position that the parser's decision lag typically extends beyond five or six words.

Closure Sentences

A 2 (early vs late closure) × 2 (long vs short length) ANOVA was performed on the total reading time per letter. Both subjects F_1 and sentences F_2 were treated as random effects in separate analyses and min F' was calculated based on the variance due to both subjects and sentences (Clark, 1973). The analyses of variance revealed that early closure sentences required more reading time than late closure sentences, $F_1(1,15) = 15.85, p < .01, F_2(1,15) = 10.16, p < .01$, and min $F'(1,29) = 6.19, p < .05$. The interaction was also significant, $F_1(1,15) = 14.94, p < .01, F_2(1,15) = 7.10, p < .05$, and min $F'(1,27) = 4.81, p < .05$. The data, as shown in Table 1, clearly indicate that the early closure–long sentences resulted in longer reading times than the other three sentence types.

In order to examine the results more precisely, separate 2 (closure) × 2 (length) × 3 (region of the sentence: prior to ambiguity, ambiguity, and disambiguating region) × 2 (pass: first vs second) ANOVAs were performed on reading time per letter for the different regions of the sentences as defined previously. These data are shown in Table 2. Reading time was

TABLE 1
Mean Reading Time per Letter (msec) for Each of the Four Closure Sentence Versions

	Early closure	Late closure	$\bar{X}$
Long	68 (176)	50 (240)	59
Short	57 (211)	55 (218)	56
$\bar{X}$	62.5	52.5	

Note. Values in parentheses represent the estimated reading rate in words per minute based on an average word length of 5 characters.

TABLE 2
Mean Reading Time per Letter in the Different Regions of the Sentence for the Different Versions of the Closure Sentences on the First and Second Pass

Sentence type	Region of the sentence		
	Before ambiguity	Ambiguity	Disambiguation
Early closure–long			
1st pass	44	40	54
2nd pass	21	32	48
Total	65	72	102
Early closure–short			
1st pass	43	37	41
2nd pass	18	37	41
Total	61	74	82
Late closure–long			
1st pass	43	35	40
2nd pass	12	15	23
Total	55	50	63
Late closure–short			
1st pass	40	42	47
2nd pass	16	27	22
Total	56	69	69

longer overall on the first pass than the second pass, $F(1,15) = 20.99, p < .001$. Consistent with the total reading time analysis, early closure sentences required more reading time than did the late closure sentences, $F(1,15) = 8.66, p < .01$. Also, consistent with the previous analysis, the interaction of Closure × Length was significant, $F(1,15) = 4.47, p < .05$.

The most important result to be gleaned from this second analysis was the fact that reading times differed rather dramatically in the three critical regions of the sentence, $F(2,30) = 11.15, p < .001$. A Newman–Keuls test indicated that the reading time per letter (first pass and second pass) in the region prior to the ambiguous phrase (60 msec per letter) did not differ from the reading time in the ambiguous region (66 msec per letter), but reading time in each of these regions did differ ($p < .01$) from reading time in the disambiguating region (80 msec per letter). Perhaps a clearer picture of the pattern of results can be obtained by examining the significant interaction of Pass × Region, $F(2,30) = 5.29, p < .05$. On the first pass, reading time in the disambiguating region was slightly longer than in the region prior to the ambiguity and considerably longer than in the ambiguous region.[3] On the second pass, reading time was again longest in the disambiguating region, but more time was spent in the ambiguous region than in the region prior to it. Thus, the general pattern that emerges from these data is that the subjects read the sentence quite smoothly up to the point of the disambiguation and from that point on relatively more time was spent in the disambiguating and ambiguous region than in the region prior to the ambiguity. We shall return to this point later when we discuss the pattern of eye movements.

There were also significant interactions of Closure × Pass, $F(1,15) = 8.33, p < .02$, and Closure × Region, $F(2,30) = 2.83, p < .05$. Although the early closure sentences resulted in longer reading times than the late closure sentences on both the first and second pass, the effect was more pronounced on the second pass. Although the early closure sentences resulted in longer reading times than the late closure sentences in all three regions, the difference was more pronounced in the disambiguating region.

As we indicated previously, the closure sentences used in the experiment were of two types. An examination of the Appendix reveals that sentences 1–8 in the closure set consisted of an initial subordinate clause followed by a single main clause. Sentences 9–16 consisted of three conjoined clauses. It may have been the case that many of our effects regarding the differences between early and late closure sentences were due to the apparently more difficult construction presented by the conjoined clauses. To ascertain the extent to which this may have been the case, we carried out an ANOVA comparing the early closure–long and late closure–long sentences in the three critical regions of the sentence for the sentences with the subordinate clause in comparison to the sentences with the conjoined clauses. As would be expected from our prior analyses, there were significant main effects of closure type, $F(1,15) = 16.9, p < .01$, and region, $F(2,30) = 4.35, p < .05$, and the interaction of Closure Type × Region, $F(2,30) = 4.53, p < .05$, was also significant. However, the main effect of clause type (subordinate vs conjoined) was not significant, $F(1,15) = 3.41, p > .08$, while the interaction of Closure Type × Clause Type was significant, $F(1,15) = 5.36, p < .05$. As seen in Table 3, the early closure sentences were more difficult than the late

[3] There was no reading time difference between the four sentence types in the ambiguous region on the first pass (35–42 msec per letter). The difference between the region prior to the ambiguous phrase and the ambiguous phrase itself is attributable to rather long initial fixations (and sometimes long second fixations as well) on the sentence (which often exceeded 400–500 msec by some subjects). Such a finding has often been reported in the literature on eye movements in reading (cf. Rayner, 1978) and occurred in both the experimental and filler sentences. When these initial long fixations were removed from the data, reading time per letter in the region prior to the ambiguity became comparable to the reading time in the region of the ambiguous phrase.

TABLE 3
Mean Reading Time per Letter in the Different Regions of the Sentence for Closure Sentences with Subordinate and Conjoined Clauses

	Before ambiguity	Ambiguity	Disambiguation	Mean
Late closure				
Subordinate clause	57	49	59	55
Conjoined clause	54	47	57	53
Mean	55.5	48	58	
Early closure				
Subordinate clause	65	66	79	70
Conjoined clause	75	79	113	89
Mean	70	72.5	96	

closure sentences for both the conjoined and subordinate clause type sentences.[4] Separate ANOVAs on the subordinate clause and conjoined clause sentences yielded significant main effects of both region, $F(2,30) = 7.01, p < .01$, and closure type $F(1,15) = 24.36, p < .001$, and a significant interaction, $F(2,30) = 5.80, p < .01$ for the conjoined sentences. For the subordinate clause sentences, the early closure sentences were more difficult than the late closure sentences, $F(1,15) = 4.85, p < .05$.

The final analyses carried out on the closure sentences dealt with average fixation duration. Our interest in this analysis was to determine the point at which the eye movement records indicated via the fixation duration that subjects were having difficulty processing the sentence. Hence, for each subject reading each of the four sentence types, we computed the mean fixation duration on the last three fixations prior to encountering the disambiguating words, as well as the duration on the next three fixations beginning with the first fixation in the disambiguating material. A 2 (closure) $\times$ 2 (length) $\times$ 6 (serial order of fixation) ANOVA indicated that average fixation durations were longer for early closure sentences than late closure, $F(1,15) = 10.5, p < .01$. The only other significant effect was for order of fixation, $F(5,75) = 4.30$, $p < .01$. A Newman–Keuls test indicated that average fixation durations were shorter ($p < .05$) on the three fixations prior to the disambiguation than on the subsequent three fixations. The data are shown in Table 4.

In order to understand the point at which the readers encountered difficulty in processing the sentences, we also compared the average fixation duration on the last fixation prior to disambiguation and the first fixation on the disambiguating material. There was no difference in the

[4] We would in fact expect a late closure error to be easier to revise in the subordinate clause sentences than in the conjoined clause sentences, since the existence of a second clause (which must contain a subject noun phrase) can be predicted in advance in the subordinate clause sentences but not in the conjoined clause sentences.

TABLE 4
Average Fixation Duration on the Three Fixations Prior to Reaching the Disambiguating Region (d) and the First Three Fixations in the Sentence Following the Initial Encounter with the Disambiguating Word

Sentence type	Serial order of fixation 1 (d − 3)	2 (d − 2)	3 (d − 1)	4 (d)	5 (d + 1)	6 (d + 2)
Early closure–long	252	259	236	301	285	313
Early closure–short	245	227	245	283	267	277
Late closure–long	248	239	243	260	247	242
Late closure–short	228	239	243	268	248	242

Note. These data were computed independent of the particular region of the sentence and consist only of the serial order that the fixations occurred in.

late closure sentences ($t < 1$ for both the long and short version). For the early closure–long sentences, the first fixation on the disambiguation was longer than the previous fixation, $t(15) = 3.72$, $p < .01$. In the case of the early closure–short sentences, although the same pattern emerged, the effect was not significant, $t(15) = 1.50, p < .10$. Thus, these data suggest that when the reader was garden-pathed there was an awareness at some level on the first fixation in the disambiguating region that something was wrong as evidenced by the longer fixation duration. As seen in Table 4, the average fixation duration remained somewhat higher than average on the next two fixations as well.

Attachment Sentences

A 2 (minimal versus nonminimal attachment) × 2 (long vs short length) ANOVA was performed on the total reading time per letter and indicated that the nonminimal attachment sentences resulted in longer reading times than the minimal attachment sentences, $F_1(1,15) = 8.20, p < .02$, and $F_2(1,15) = 5.21, p < .05$. The interaction was also significant, $F_1(1,15) = 11.7, p < .01$, and $F_2(1,15) = 5.31, p < .05$. The data are shown in Table 5. Although F_1 and F_2 were both significant, the results did not quite reach ($.05 < p < .10$) statistical significance with the very conservative min F test. However, it should be noted that the pattern of results was identical to that found with the closure sentences and that the quasi-F ratio employed is ultraconservative. Thus, it is quite clear that the nonminimal attachment sentences were more difficult to read than the minimal attachment sentences and that the long version of the nonminimal attachment sentences were more difficult to process than the other versions.

As with the closure sentences, we also performed analyses on the reading time per letter in different regions of the sentence. However, it is the case with the minimal attachment versions of the sentence that there

TABLE 5
Reading Time per Letter (msec) for Each of the Four Attachment Sentence Versions

	Nonminimal attachment	Minimal attachment	$\bar{X}$
Long	61 (197)	45 (270)	53
Short	51 (235)	49 (246)	50
$\bar{X}$	56	47	

Note. Values in parentheses represent the estimated reading rate in words per minute based on an average word length of 5 characters.

was no disambiguating region in the sentence; the last word of the ambiguous noun phrase was also (usually) the last word in the sentence. Hence, we performed two separate analyses concerning the different regions of the sentence. In the first analysis, a 2 (attachment) × 2 (length) × 2 (region of the sentence: prior to ambiguity vs ambiguity) × 2 (pass) ANOVA was performed. The data are shown in Table 6. Reading time was longer on the first pass than the second pass, $F(1,15) = 185.61, p < .001$. Consistent with the total reading time, the nonminimal attachment sentences resulted in longer reading time than the minimal attachment sentences, $F(1,15) = 4.83, p < .05$ and the Attachment × Length interaction was significant, $F(1,15) = 10.37, p < .01$.

TABLE 6
Mean Reading Time per Letter in the Different Regions of the Sentence for the Different Versions of the Attachment Sentences on the First and Second Pass

	Region of the sentence		
Sentence type	Before ambiguity	Ambiguity	Disambiguation
Nonminimal attachment–long			
1st pass	43	37	51
2nd pass	17	22	30
Total	60	59	81
Nonminimal attachment–short			
1st pass	43	36	47
2nd pass	10	15	23
Total	53	51	70
Minimal attachment–long			
1st pass	41	36	--
2nd pass	7	7	--
Total	48	43	--
Minimal attachment–short			
1st pass	42	36	--
2nd pass	8	12	--
Total	50	48	--

The important result to be gleaned from this analysis was that reading times did not differ in the two regions of the sentence under consideration ($F < 1$). However, the interaction of Pass × Region was significant, $F(1,15) = 11.95, p < .01$, and is generally attributable to more time in the ambiguous region than in the region prior to the ambiguity on the second pass coupled with less overall time in the ambiguous region on the first pass. Finally, the Attachment × Pass interaction was significant, $F(1,15) = 4.88, p < .05$, as was the Attachment × Length interaction, $F(1,15) = 10.37$, $p < .01$. The former interaction was due to relatively more time on the second pass for the nonminimal attachment sentences combined with approximately the same amount of time per sentence type on the first pass while the latter interaction was due to more time per letter for the long version of nonminimal attachment than the short version coupled with the opposite trend for the minimal attachment version.

The second analysis performed on the reading time per letter excluded the minimal attachment data so that a comparison could be made of the different regions of the sentence including the disambiguating region. A 2 (length) × 3 (region of the sentence: prior to the ambiguity, ambiguity, and disambiguating) × 2 (pass) ANOVA again indicated that more time was spent per character on the first pass than the second pass, $F(1,15) = 50.52$, $p < .001$. In this analysis, the longer version resulted in longer reading times than the shorter version, $F(1,15) = 7.15$, $p < .05$. Most importantly, there was a significant effect of region, $F(2,30) = 14.88, p < .001$. A Newman–Keuls test indicated that more time was spent per character in the disambiguating region (76 msec) than in either the region prior to the ambiguity (57 msec) or the region of the ambiguity (55 msec). In general, the results of the two analyses performed on the reading time per letter data indicate that in the region prior to the ambiguity the reading time did not differ for the four versions of the sentences. Once again, there was some indication that reading times were slightly longer in the region prior to the ambiguity than in the ambiguous region on the first pass. As with the closure sentences, this difference is mainly attributable to rather long initial fixations on the sentences. Relatively speaking, on the second pass there was more time spent on the ambiguous region than the region prior to the ambiguity. In the case where there was a disambiguating region, more total time was spent in that region than in either of the other two regions of the sentence.

The final analysis performed on the attachment sentences concerned the average fixation duration in the nonminimal attachment versions for the last three fixations prior to the first fixation on the disambiguating region and the three subsequent fixations. A 2 (length) × 6 (serial order of fixation) ANOVA yielded only a significant effect for order of fixation, $F(5,75) = 2.92, p < .02$. A Newman–Keuls test indicated that average fixation durations were shorter ($p < .05$) on the three fixations prior to the disambiguation than on the three subsequent fixations. The data are shown in Table 7. As with the closure sentences, we compared the average fixation duration on the last fixation prior to the disambiguation and the first fixation on the disambiguating material. For both the long and short version, the last fixation prior to the disambiguation was significantly shorter than the first fixation on the disambiguating material, $t(15) = 2.17$ and $2.30, p < .05$ for the long and short versions, respectively. Once again, the fixation duration data reveal difficulty processing the sentence on the first fixation in the disambiguating region when the subjects were garden-pathed.

Pattern of Eye Movements

In addition to the analyses that have been described above, we also examined the pattern of eye movements that occurred when subjects were garden-pathed as they read the experimental sentences. This analysis was by nature more subjective than the preceding analyses, but the results were rather informative. In those instances where it was clear that the subject had misparsed the sentence (e.g., particularly in the long versions of early closure and nonminimal attachment sentences), three or four patterns of eye movement behavior occurred which we shall attempt to characterize. In some cases, subjects read the ambiguous noun phrase and upon reading the disambiguating region made very long fixations. These long fixations were also accompanied by very short saccades (whereas the average saccade length had been between 7 and 8 character spaces on average prior to the disambiguation, they averaged 2–3 character spaces). It was very clear that the reader was having difficulty understanding the sentence, but the eye movements generally continued in a forward direction through the sentence. Upon reading the end of the sentence, the subject then made a long regression to the beginning of the sentence and reread the sentence. The long fixations and short saccades

TABLE 7
Average Fixation Duration on the Three Fixations Prior to Reaching the Disambiguating Region (d) and the First Three Fixations in the Sentence Following the Initial Encounter with the Disambiguating Word

	Serial order of fixation					
Sentence type	1 (d − 3)	2 (d − 2)	3 (d − 1)	4 (d)	5 (d + 1)	6 (d + 2)
Nonminimal attachment–long	248	259	258	291	284	301
Nonminimal attachment–short	247	235	226	292	280	267

Note. These data were computed independent of the particular region of the sentence and consist only of the serial order that the fixations occurred in.

in the disambiguating region and thereafter may also have been accompanied by short regressions, but the reader did not regress at that point back to the beginning of the sentence or to the ambiguous region. We shall characterize this behavior as *chaos* in that the reader apparently was having great difficulties understanding the sentence but seemed to have no insights as to what the nature of the processing difficulties were. This pattern of eye movements was particularly noticeable among three of the subjects and occurred less frequently with most of the other subjects.

We will characterize the other major type of eye movement behavior when the reader misparsed the sentence as *disruption*. In these instances, the eye movement pattern took a number of forms. In some cases, upon encountering the disambiguating word, the subject simply made an unusually long fixation (e.g., exceeding 300–400 msec) followed by one or two other long fixations (and often short saccades, as well). It appears that in these cases, the reader was able to reanalyze the sentence and resolve the ambiguity, but it took some additional processing after encountering the disambiguation. In other cases, the reader fixated on the disambiguating region for an average amount of time and then immediately made a regression back to the ambiguous region of the sentence. In a similar manner, upon encountering the disambiguating word, subjects often made a very long fixation followed by a regression back to the ambiguous region. It should be noted that these patterns characterized as part of the category we have called *disruption* may be a bit of an idealization and that sometimes there was more than one fixation and the increased fixation duration may not always have been on the first fixation in the disambiguating region; but the point is that in these cases subjects did show longer than average fixations and often made regressions to the ambiguous region. In order to quantify these notions somewhat, we computed all of the regressions made in the experimental sentences as a function of the region that the regression was initiated from and the region of the sentence the eye landed on following the regression. These data are shown in Table 8. Regressions for all eight sentence types are included in this analysis, so that the odds are loaded against the position that we have argued above. However, the values in parentheses are the proportions of regressions for only the early closure and nonminimal attachment sentences. Of all the regressions, 70% were in closure sentences. Seventy-one percent of the regressions were from early closure and nonminimal attachment sentences (and 57% of these regressions were from the long versions). It can be seen quite clearly in Table 8 that about half the regressions made were to the ambiguous region and about one-third of the regressions were initiated from the disambiguating region with the eye landing in the ambiguous region. Of the regressions ending in the ambiguous region, 20% of them ended on the first word of the ambiguous phrase.

TABLE 8

Matrix of the Region of the Sentence that a Regression was Initiated from and the Region of the Sentence that the Eye Movement Ended in

Regression Initiated from	Eye Movement Ended				
	Beginning of sentence	Before ambiguity	In ambiguity	In disambiguation	Total
Disambiguating region[b]	.03 (.01)[a]	.04 (.03)	.33 (.36)	.12 (.11)	.52 (.51)
After disambiguating region	.01 (.02)	.04 (.03)	.06 (.05)	.04 (.05)	.15 (.15)
End of sentence	.18 (.17)	.03 (.03)	.10 (.12)	.02 (.02)	.33 (.34)
Total	.22 (.20)	.11 (.09)	.49 (.53)	.18 (.18)	

Note. The values are proportions and are based on 222 regressions made by the subjects. The range of regressions was 5–29 per subject.

[a] The values in parentheses are for the early closure and nonminimal attachment sentences.

[b] In the minimal attachment sentences, the end of the sentence was considered to be the disambiguating region.

Regressions for the two types of closure sentences are presented separately in Table 9 and regressions for the two types of attachment sentences are presented in Table 10.

It should be noted that none of the tables includes a category for eye movements ending in the region following the disambiguation. This is because subjects did not regress to that region; they either made a regression to the ambiguous region or to the beginning of the sentence. Generally, there was no indication of a series of right-to-left eye movements back through the sentence. In a very few cases, there was a single intervening fixation between the point where the regression was initiated and the ambiguous region or the beginning of the sentence. However, this occurred only in long sentences which extended over multiple lines. The duration of the fixation was short and was never followed by a rightward fixation which was itself followed by a continuation of the leftward regression. Thus, there was no evidence for backward reanalysis which would be indicated by a series of systematic right-to-left eye movements through the sentence.

DISCUSSION

The overall view of sentence processing suggested by the results is one in which the parser immediately assigns local structure to fixated items and notes any structural incompatibility of these items and the analysis of preceding material. The results did not provide any indication that temporary phrase structure ambiguities are detected: rather, the parser initially assigns just the preferred (e.g., late closure or minimal attachment) structure to the ambiguous material and then, later, revises this structure if it should prove untenable. The parser's revision procedures appear able to quickly identify the source of an erroneous analysis of temporarily ambiguous material, often permitting the parser to revise just that part of

TABLE 9
Matrix of the Region of the Sentence that a Regression Was Initiated from and the Region of the Sentence that the Eye Movement Ended in for Closure Sentences

Regression initiated from	Eye Movement Ended: Beginning of sentence E	L	Before ambiguity E	L	In ambiguity E	L	In disambiguation E	L	Total E	L
Disambiguating region	.00	.04	.03	.00	.38	.30	.10	.16	.51	.50
After disambiguating region	.02	.00	.02	.02	.09	.10	.02	.02	.16	.14
End of sentence	.14	.21	.02	.04	.15	.10	.02	.02	.33	.37
Total	.16	.25	.07	.06	.62	.50	.14	.20		

Note. E, early closure (70% of the regressions were in early closure sentences); L, late closure.

TABLE 10
Matrix of the Region of the Sentence that a Regression Was Initiated from and the Region of the Sentence that the Eye Movement Ended in for Attachment Sentences

Regression initiated from	Eye movement ended: Beginning of sentence N	M	Before ambiguity N	M	In ambiguity N	M	In disambiguation N	M	Total N	M
Disambiguating region	.04	.20	.02	.09	.33	.15	.13	.09	.52	.53
After disambiguating region	.00	.00	.04	.15	.00	.09	.09	.00	.13	.24
End of sentence	.22	.15	.04	.04	.07	.04	.02	.00	.35	.23
Total	.26	.35	.10	.28	.40	.28	.24	.09		

Note. N, nonminimal attachment (72% of the regressions were in nonminimal attachment sentences); M, minimal attachment.

its initial analysis which resulted in the conflict with subsequent disambiguating material. Occasionally, however, the parser will plow on through the sentence after incompatible information is received, apparently without revising its initial analysis, as if optimistically hoping that subsequent context would aid in its attempts to resolve the discrepancy with the disambiguating material.

Specifically, the data confirm the predictions of the garden-path theory of comprehension, together with the particular strategies tested. The data also support the predictions of the selective reanalysis hypothesis, over the forward or backward reanalysis hypotheses. We turn now to a more detailed discussion of the evidence for each of these hypotheses.

The Garden-Path Theory

In contrast to the parallel processing hypothesis and the minimal commitment hypothesis, the garden-path theory predicted that the reading time for the different experimental sentences should depend on whether the material following the ambiguity was compatible with the chosen analysis of the ambiguous phrase. The fact that reading times were significantly longer for the early closure and nonminimal attachment sentences than for the late closure and minimal attachment sentences and that the difference in reading times was associated with the disambiguating region of the sentence clearly supports this prediction. Further, the fact that reading time and average fixation duration were not longer in the ambiguous region of the sentence than in the (unambiguous) region prior to the ambiguity argues against the view that phrase structure ambiguities are detected or that more than one analysis of the ambiguous material is computed on the parser's first pass through the sentence. This finding, together with the finding that average fixation durations are longer in the disambiguating region of early closure and nonminimal attachment sentences than in earlier regions (while this is not the case in the late closure and minimal attachment sentences) also provides evidence that the parser analyses ambiguous material in advance of disambiguating context, contrary to the minimal commitment hypothesis. Finally, these results support the late closure and minimal attachment strategies which were used to predict which analysis of the ambiguous phrase would initially be pursued.

One question to be addressed is whether subjects were ever garden-pathed in the late closure or minimal attachment sentences. Two nonsignificant tendencies in the data suggest that, at least in the late closure sentences, subjects may occasionally have been garden-pathed: there is a slight increase in the duration of the first fixation in the disambiguating region of the late closure sentences (Table 4); reading times are also somewhat longer in the disambiguating region of the late closure sentences than in the earlier parts of the sentence (Table 2). Given that the first word in the disambiguating region of the late closure sentences was always the first word of a clause, these tendencies might simply be due to additional processing operations associated with the opening of a new clause (e.g., completing the semantic interpretation of the preceding clause). Another possibility, however, is that some of the controls used in the construction of the experimental sentences may not have been totally effective and thus a few of the sentences may have temporarily permitted an unforseen analysis. In particular, subjects seemed to have problems processing *all* versions of the sentence *Wherever Alice walks her dog men follow* (i.e., including the late closure versions). Here we expected the anomalous nature of the noun phrase *her dog men* to exclude the (super-late closure) analysis of the sentence, where *men* is analyzed as forming a constituent with the preceding items *her* and *dog*. The data suggest that this particular control was not effective.

In the minimal attachment sentences, the ambiguous phrase typically ended the sentence and thus it was only the absence of subsequent material (i.e., the period) which would disambiguate these sentences. Thus, if subjects were garden-pathed in the minimal attachment sentences, this should have resulted in longer reading times in the ambiguous region of these sentences than in the region before the ambiguity. However, there was not even a nonsignificant tendency in this direction (Table 6).

Another question is whether subjects were garden-pathed in sentences with short ambiguous phrases. If the parser delays decisions about how to structure an item with the other phrases of the sentence until one or two words subsequent to that item have been fixated, then in the sentences with short ambiguous phrases subjects would have fixated the words of the disambiguating region prior to making a decision about the preceding ambiguous noun phrase. Use of information from disambiguating words would prevent garden-pathing, explaining the interaction of closure type or attachment type with the length of the ambiguous phrase.

This decision lag hypothesis suffers from a number of inadequacies in terms of accounting for our data. First, the different measures used in the experiment consistently indicate that the reading times and fixation durations in the disambiguating region of the early closure and nonminimal attachment sentences with short ambiguous phrases are longer than those in the earlier regions of these sentences, and they also indicate that the short versions of the early closure and nonminimal attachment sentences are more difficult to process than the short version of the late closure and minimal attachment sentences. Second, the finding that fixation durations are significantly longer on the very first fixation in the disambiguating region of the early closure and nonminimal attachment sentences is most naturally interpreted as indicating that as soon as the parser fixates the material in the disambiguating region, it tries to incorporate the disambiguating material into its analysis of preceding material and discovers that the two are incompatible. But clearly this presupposes that the parser does not delay its decisions about how to structure the fixated items with the other constituents of the sentence. Thus, this finding also runs counter to any decision lag hypothesis which claims that the structural analysis of an item is delayed until some constant number of subsequent words has been fixated.

Given the above problems with the decision lag hypothesis, we might consider whether there is some alternative explanation of the interaction of sentence type and the length of the ambiguous phrase. One alternative is that the parser's decision lag is variable and not under control of linguistic factors. A more interesting alternative is available if we assume that the higher level (i.e., post-lexical) semantic interpretation of linguistic material lags slightly behind its syntactic analysis. The observed interaction might then be explained by assuming that revising processing errors which have already affected the semantic interpretation of a sentence is more costly than revising errors which can be corrected before they have any effect on the semantic processing of the sentence. Subjects may have been syntactically garden-pathed in both the short and the long version of the early closure and nonminimal attachment sentences. However, in the sentences with short ambiguous phrases, the erroneous analysis may have been revised before it affected the semantic interpretation of the sentence; whereas, in sentences with long ambiguous phrases, the syntactic misanalysis of the ambiguous phrase may have had semantic consequences before the disambiguating material was encountered. This hypothesis would account for our findings and it does not suffer from the problems of the (constant size) decision lag hypothesis nor from the near vacuity of the variable (random) decision lag hypothesis. One final possibility is that revising an analysis of the last one or two items fixated is relatively cost-free simply because of the salience of these items in memory.

Selective Reanalysis

We will begin discussion of reanalysis procedures by briefly summarizing the different patterns of recovery from a garden path indicated by our results. When subjects fixated the disambiguating region of a sentence, disruption in their processing of the sentence was indicated by (a) longer fixation durations (not necessarily followed by a regressive eye movement), (b) a regressive eye movement returning to an earlier portion of the disambiguating region of the sentence, or (c) a regressive eye movement returning to the ambiguous phrase. If subjects made a regressive eye movement which was initiated from one of the regions of the sentence following the disambiguating region, the eye movement typically ended in the ambiguous region (.16 of all regressions) or at the beginning of the sentence (.19 of all regressions; only .01 of these regressions were initiated from before the very end of the sentence).

Longer fixation durations in the disambiguating region which were not followed by regressions suggest that it is not absolutely necessary for subjects to fixate ambiguous material in order to revise an inappropriate analysis of it. Regressions from the disambiguating region to an earlier part of the disambiguating region may indicate that subjects want to confirm their perceptual encoding of the disambiguating material to be sure that this material does present a problem for their initial analysis of the constitutent structure of the sentence before trying to revise the analysis. Regressive eye movements from the disambiguating region to the ambiguous region (the dominant type of regression) probably indicate that subjects have detected an error in their initial analysis of the sentence and have identified the source of the error. The fact that one-third of the regressions initiated from after the disambiguating region or from the end of the sentence ended in the ambiguous region, without any indication of intervening fixations of sufficient duration to permit reanalysis of the intervening material, argues that subjects were not automatically backtracking through the sentence. The only time subjects were likely to regress to the beginning of the sentence was when the regression was initiated from the very end of the sentence.

Turning to the reanalysis hypotheses outlined above, it is clear that the results obtained in the experiment are inconsistent with both the forward and backward reanalysis hypotheses. The forward reanalysis hypothesis predicts that all regressive eye movements should have returned to the very beginning of the sentence; the data very clearly indicate that subjects only regress to the beginning of a sentence when the regression is initiated from the very end of the sentence.

The backward reanalysis hypothesis predicts that, regardless of the point where the regression was initiated, regressive eye movements should proceed backward through the sentence, with alternative options being considered at choice points. The fact that regressions from regions after the disambiguating region did not provide any evidence for systematic backtracking or for the checking of alternative analyses in any region of the sentence other than the ambiguous region argues against the backward reanalysis hypothesis. This evidence against backward reanalysis is especially compelling given that some of the experimental sentences were very long and thus the amount of material intervening between the initiation point and the final resting point of many regressive eye movements clearly did contain numerous points where, in principle, alternative decisions could have been made (regardless of the level of structure used to define the notion "choice point"). We can not exclude the possibility that a more elaborate backward reanalysis hypothesis might be developed which would account for the data.

The selective reanalysis hypothesis predicts that eye movements should regress from the disambiguating region to the ambiguous region of the sentence, and this was in fact the dominant pattern of regressions. As noted above, it is not obvious how regressions to the beginning of the sentence should be interpreted. Such regressions may well indicate that subjects did not successfully correct their erroneous analysis of the sentence during their first pass through it. This would provide evidence that selective reanalysis is not always possible, even in sentences which do contain a sufficient amount of useable information to permit reanalysis of just that portion of the sentence that gave rise to the error signal.

We have argued that the parser's dominant mode of correcting processing errors is best described by the selective reanalysis hypothesis. In the introduction, we noted that intuitive evidence suggests that there are differences both in the ease of reanalyzing different specific sentences and differences in the complexity associated with recovering from distinct types of garden paths. The selective reanalysis hypothesis provides a framework for explaining why such differences should exist. However, the hypothesis is obviously in need of considerable refinement if it is to provide a detailed account and explanation of the precise nature of these differences. To flesh out the selective reanalysis hypothesis, presumably it will be necessary to determine what counts as a perspicuous signal that a particular type of error has occurred and to explicitly characterize the various properties that render an error difficult to revise (e.g., changes in the relation of already identified phrases seem easier than changes involving the identification of new phrases; errors involving a "lower level" of analysis seem extremely difficult to revise perhaps because the processor is no longer concerned with that level of analysis or perhaps because the error will have had a chance to propagate through subsequent stages of processing, etc.). Further, we think it is important to distinguish information which affects the first pass analysis of a sentence from information which affects reanalysis.

CONCLUSIONS

The recording of eye movements during the comprehension of natural language sentences provides a valuable technique for measuring the on-line processing complexity of linguistic material, a technique sensitive to the parser's structural analysis of sentences. The results of the present experiment (as well as others, e.g., Rayner, 1977; Rayner, 1978; McConkie et al., 1979; Just & Carpenter, 1980; Carpenter & Daneman, 1981) indicate that there is a rather tight correlation between the position and duration of fixations and the processing operations involved in the comprehension of the material being read. The technique permits various aspects of sentence comprehension to be examined which would be difficult if not impossible to explore using end-of-sentence measures of processing complexity (e.g., the question of whether structural ambiguities are detected at some subconscious level of awareness, questions concerning the *timing* of the parser's decisions, when some analysis of a sentence is discovered to be inappropriate, etc.).

We have used eye movement data to argue that the parser pursues just a single analysis of structurally ambiguous sentences, detecting the ambiguity only if the chosen analysis of it should happen to prove incompatible with subsequent material. In the present experiment we only examined cases where there was a gross incompatibility which would result in an ungrammaticality if the parser's initial analysis were left unrevised; clearly more subtle incompatibilities should also be investigated. Typically the incompatibility between the disambiguating material and the analysis of preceding material was detected as soon as the disambiguating material was fixated (as indicated by longer fixation durations).

Results supported the garden-path theory of sentence comprehension, along with the late closure and minimal attachment strategies. An interaction of sentence type and the length of the ambiguous phrase was taken to suggest that errors in the syntactic analysis of a sentence are more costly to revise if there has been sufficient time for the error to affect the semantic processing of the sentence.

Finally, the results supported the selective reanalysis hypothesis which, in turn, offers an explanation for why there should be differences in the processing complexity associated with recovering from different garden paths. Ultimately, the selective reanalysis hypothesis promises both to provide an explanation for why some garden paths are consciously detected, while others are not, and to contribute to a more adequate characterization of the class of unparsable sentence structures.

APPENDIX: CLOSURE AND ATTACHMENT SENTENCES USED IN THE EXPERIMENT

Closure Sentences: (a) Late Closure Long
(b) Early Closure Long
(c) Late Closure Short
(d) Early Closure Short

1a. Though George kept on reading that stupid science fiction story Susan bothered him.
1b. Though George kept on reading that stupid science fiction story really bothered him.
1c. Though George kept on reading the story Susan bothered him.
1d. Though George kept on reading the story really bothered him.

2a. After you drank the strange looking water they discovered it was polluted.
2b. After you drank the strange looking water was discovered to be polluted.
2c. After you drank the water they discovered it was polluted.
2d. After you drank the water was discovered to be polluted.

3a. While Mary was mending the old grandfather clock in the hall it started chiming.
3b. While Mary was mending the old grandfather clock in the hall started to chime.
3c. While Mary was mending the clock it started chiming.
3d. While Mary was mending the clock started to chime.

4a. Before the king rides his beautiful white horse it's always groomed.
4b. Before the king rides his beautiful white horse is always groomed.
4c. Before the king rides his horse it's always groomed.
4d. Before the king rides his horse is always groomed.

5a. Wherever Alice walks her shaggy sheep dog men follow.
5b. Wherever Alice walks her shaggy sheep dog will follow.
5c. Wherever Alice walks her dog men follow.
5d. Wherever Alice walks her dog will follow.

6a. Because my son likes to visit people who are older the neighbors think he's terrific.
6b. Because my son likes to visit people who are older think he's a terrific neighbor.
6c. Because my son likes to visit old people the neighbors think he's terrific.
6d. Because my son likes to visit old people think he's a terrific neighbor.

7a. Since Jay always jogs a mile and a half this seems like a short distance to him.
7b. Since Jay always jogs a mile and a half really seems like a very short distance to him.
7c. Since Jay always jogs a mile this seems like a short distance to him.
7d. Since Jay always jogs a mile really seems like a very short distance to him.

8a. Though Hilda finally agreed to sing the German Christmas carols it turned out her voice was just awful.
8b. Though Hilda finally agreed to sing the German Christmas carols she chose turned out to be just awful.
8c. Though Hilda finally agreed to sing the songs it turned out her voice was just awful.
8d. Though Hilda finally agreed to sing the songs she chose turned out to be just awful.

9a. Anne was watching the people on the street she laughed and nobody knew why.
9b. Anne was watching the people on the street were laughing and nobody knew why.
9c. Anne was watching you she laughed and nobody knew why.
9d. Anne was watching you were laughing and nobody knew why.

10a. John was hitting the little blonde boy Gerry was cheering and Tom was shouting.
10b. John was hitting the little blonde boy was pitching and Gerry was coaching Tom.
10c. John was hitting Jack Gerry was cheering and Tom was shouting.
10d. John was hitting Jack was pitching and Gerry was coaching Tom.

11a. The children were playing a very comical game of football the television was on and Granny was knitting.
11b. The children were playing a very comical game of football was on television and Granny was knitting.
11c. The children were playing football the television was on and Granny was knitting.
11d. The children were playing football was on television and Granny was knitting.

12a. The irate customers were cursing the incompetent young manager all the waiters were running back and forth and everyone seemed extremely upset.
12b. The irate customers were cursing the incompetent young manager was running back and forth to the kitchen and everyone seemed extremely upset.
12c. The irate customers were cursing the manager all the waiters were running back and forth and everyone seemed extremely upset.
12d. The irate customers were cursing the manager was running back and forth to the kitchen and everyone seemed extremely upset.

13a. Last year my roommate was constantly cleaning our little farm house it always looked absolutely spotless and I couldn't stand it.
13b. Last year my roommate was constantly cleaning our little farm house always looked absolutely spotless and I simply couldn't stand it.
13c. Last year my roommate was constantly cleaning the house it always looked absolutely spotless and I couldn't stand it.
13d. Last year my roommate was constantly cleaning the house always looked absolutely spotless and I simply couldn't stand it.

14a. Sally spent all last night writing a long term paper her history exam is tomorrow and she looks completely exhausted.
14b. Sally spent all last night writing a long term paper will be due tomorrow and she already looks completely exhausted.
14c. Sally spent all last night writing another paper her history exam is tomorrow and she looks completely exhausted.
14d. Sally spent all last night writing another paper will be due tomorrow and she already looks completely exhausted.

15a. My little brother is cooking the barbecued chicken wings everything else is burned to a crisp and so apparently we won't have anything to eat for dinner.
15b. My little brother is cooking the barbecued chicken wings are burned to a crisp and so apparently we're not going to have anything to eat for dinner.
15c. My little brother is cooking the chicken everything else is burned to a crisp and so apparently we won't have anything to eat for dinner.
15d. My little brother is cooking the chicken is burned to a crisp and so apparently we're not going to have anything to eat for dinner.

16a. Susie was pushing the big red wagon it wouldn't move an inch and so the boys were making fun of her.

16b. Susie was pushing the big red wagon wouldn't move an inch and so the boys were all making fun of her.
16c. Susie was pushing the wagon it wouldn't move an inch and so the boys were making fun of her.
16d. Susie was pushing the wagon wouldn't move an inch and so the boys were all making fun of her.

Attachment Sentences: (a) Minimal Attachment Long
(b) Non-minimal Attachment Long
(c) Minimal Attachment Short
(d) Non-minimal Attachment Short

1a. Sally was relieved when she found out the answer to the difficult physics problem.
1b. Sally found out the answer to the difficult physics problem was in the book.
1c. Sally was relieved when she found out the answer.
1d. Sally found out the answer was in the book.

2a. It's possible the detective didn't see the man with a gun in his hand.
2b. The detective probably saw the man with a gun in his hand fall over.
2c. It's possible the detective didn't see the man.
2d. The detective probably saw the man fall over.

3a. I wonder if Tom heard the latest gossip about the new neighbors.
3b. Tom heard the latest gossip about the new neighbors wasn't true.
3c. I wonder if Tom heard the gossip.
3d. Tom heard the gossip wasn't true.

4a. It appears that Sherlock Holmes didn't suspect the very beautiful young countess.
4b. Sherlock Holmes didn't suspect the very beautiful young countess was a fraud.
4c. It appears that Sherlock Holmes didn't suspect the countess.
4d. Sherlock Holmes didn't suspect the countess was a fraud.

5a. I suppose it's possible that the clerk knows the woman wearing that outrageous peacock hat.
5b. I suppose that the clerk knows the woman wearing that outrageous peacock hat is crazy.
5c. I suppose it's possible that the clerk knows those women.
5d. I suppose that the clerk knows those women are crazy.

6a. Nobody knew why the speaker concluded his very interesting but technical lecture with such haste.
6b. Apparently the speaker concluded his very interesting but technical lecture had not been a success.
6c. Nobody knew why the speaker concluded his lecture with such haste.
6d. Apparently the speaker concluded his lecture had not been a success.

7a. The commissioner suspects that the financial committee failed to mention the very large accounting error.
7b. The financial committee failed to mention the very large accounting error was their own fault.
7c. The commissioner suspects that the financial committee failed to mention the error.
7d. The financial committee failed to mention the error was their own fault.

8a. They say the city council argued the radical young mayor's position forcefully.
8b. The new city council argued the radical young mayor's position was immoral.

8c. They say the city council argued their position forcefully.
8d. The new city council argued their position was immoral.

9a. The lawyers think his second wife will claim the entire family inheritance.
9b. His second wife will claim the entire family inheritance belongs to her.
9c. The lawyers think his second wife will claim the inheritance.
9d. His second wife will claim the inheritance belongs to her.

10a. John agreed that politicians typically don't explicitly announce their plans and programs.
10b. Politicians typically don't explicitly announce their plans and programs are completely unrealistic.
10c. John agreed that politicians typically don't explicitly announce their programs.
10d. Politicians typically don't explicitly announce their programs are completely unrealistic.

11a. After she'd had dozens of tests I think Julia finally believed the unconventional young doctor's diagnosis.
11b. After she'd had dozens of tests Julia finally believed the unconventional young doctor's diagnosis was accurate.
11c. After she'd had dozens of tests I think Julia finally believed the diagnosis.
11d. After she'd had dozens of tests Julia finally believed the diagnosis was accurate.

12a. We were surprised that Sam remembered our fiftieth wedding anniversary.
12b. Sam remembered our fiftieth wedding anniversary would be quite soon.
12c. We were surprised that Sam remembered our anniversary.
12d. Sam remembered our anniversary would be quite soon.

13a. We figured that Tom probably forgot most of his camping equipment.
13b. Tom probably forgot most of his camping equipment had been stolen.
13c. We figured that Tom probably forgot the flashlight.
13d. Tom probably forgot the flashlight had been stolen.

14a. Nobody realized that the policeman immediately recognized all of the people in the car.
14b. The policeman immediately recognized all of the people in the car were completely drunk.
14c. Nobody realized that the policeman immediately recognized the thieves.
14d. The policeman immediately recognized the thieves were completely drunk.

15a. Everyone thinks the democrats first suggested a new health insurance program.
15b. The democrats first suggested a new health insurance program was needed.
15c. Everyone thinks the democrats first suggested this program.
15d. The democrats first suggested this program was needed.

16a. James was pleased that the press reported the entire sordid affair.
16b. The press reported the entire sordid affair began as a prank.
16c. James was pleased that the press reported the incident.
16d. The press reported the incident began as a prank.

REFERENCES

Carpenter, P. A., & Daneman, M. Lexical access and error recovery in reading: A model based on eye fixations. *Journal of Verbal Learning and Verbal Behavior*, 1981, **20**, 137–160.

Clark, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 335–359.

Cornsweet, T. N., & Crane, H. D. Accurate two-dimensional eye tracker using first and fourth Purkinje images. *Journal of the Optical Society of America*, 1973, **63**, 921–930.

Fodor, J. A., Bever, T. G., & Garrett, M. F. *The psychology of language*. New York: McGraw-Hill, 1974.

Fodor, J. D., & Frazier, L. Is the human sentence parsing mechanism an ATN? *Cognition*, 1980, **8**, 417–459.

Frazier, L. On comprehending sentences: Syntactic parsing strategies. Unpublished doctoral dissertation, University of Connecticut, 1978. (Available from IU Linguistics Club, 310 Lindley Hall, University of Indiana, Bloomington, Indiana.)

Frazier, L., & Fodor, J. D. The sausage machine: A new two-stage model of the parser. *Cognition*, 1978, **6**, 291–325.

Just, M. A., & Carpenter, P. A. A theory of reading: From eye fixations to comprehension. *Psychological Review*, 1980, **87**, 329–354.

Kaplan, R. Augmented transition networks as psychological models of sentence comprehension. *Artificial Intelligence*, 1972, **3**, 77–100.

Kimball, J. Seven principles of surface structure parsing in natural language. *Cognition*, 1973, **2**, 15–47.

Marcus, M. *A theory of syntactic recognition for natural language*. Cambridge, Mass: MIT Press, 1980.

McConkie, G. W., Hogaboam, T. W., Wolverton, G. S., Zola, D., & Lucas, P. A. Toward the use of eye movements in the study of language processing. *Discourse Processes*, 1979, **2**, 157–177.

Rayner, K. Visual attention in reading: Eye movements reflect cognitive processes. *Memory & Cognition*, 1977, **4**, 443–448.

Rayner, K. Eye movements in reading and information processing. *Psychological Bulletin*, 1978, **85**, 618–660.

Winograd, T. Understanding natural language. *Cognitive Psychology*, 1972, **3**, 1–191.

(Accepted July 16, 1981)