

Plausibility and argument structure in sentence comprehension

SHARI R. SPEER

University of Kansas, Lawrence, Kansas

and

CHARLES CLIFTON, JR.

University of Massachusetts, Amherst, Massachusetts

In two experiments, we investigated how reading time was affected by the plausibility of the prepositional phrase in subject-verb-noun-phrase-prepositional-phrase sentences, and the status of the prepositional phrase as argument versus adjunct of the verb. Highly plausible prepositional phrases were read faster than less plausible ones, and argument prepositional phrases were read faster than adjuncts. These effects appeared both in a self-paced reading experiment and in an experiment that measured eye movements during normal reading. The effects of plausibility were substantially larger and longer lasting than the effects of argument status, but both appeared very early in the reading of the prepositional phrase. The implications of these effects for models of parsing and sentence interpretation are discussed.

In the present paper, we examine the reading of sentences such as *The political organization tried to inform the voters about the ballot issues* and *The political organization tried to inform the voters about a minute too late*. The prepositional phrases (PPs) that end these sentences differ in at least two important ways. First, the PP of the first sentence is an argument of that sentence's verb *inform*, whereas the PP of the second sentence is an adjunct or modifier of the verb phrase (VP). Second, the PP of the first sentence is notably more plausible and sensible than the PP of the second.

The initial focus of the experiments we report is on the difference between arguments and adjuncts. Argument status plays a central role in some recent theories of parsing (e.g., Abney, 1989; Crocker, 1994; Pritchett, 1992), as will be discussed below. For present purposes, a phrase is an argument of a given word if the lexical information for that word specifies the syntactic and semantic relationship between it and the phrase. For instance, the direct object of a transitive verb is an argument of the verb; the

verb's lexical entry specifies that it can take a noun phrase (NP) as a direct object, and it assigns some thematic role to the NP. As another example, a verb such as *put* has an obligatory PP as its argument, as in *I put the box on the table*. Adjuncts contrast with arguments. A phrase is an adjunct (generally, a modifier) of some other phrase if the syntactic and semantic relation of the adjunct to the word that heads the phrase it modifies is not specific to that lexical item, but rather holds generally for instances of the same part of speech. A clear case of an adjunct is a PP specifying the time or manner of some action: *I ate the meal in the morning* or *I ate the meal in a hurry*. The structural difference between arguments and adjuncts can be illustrated by the tree structures in Figure 1. In the top panel, the argument *on the table* is represented as the sister of its argument assigner, the verb. In the bottom panel, the adjunct *in a hurry* is represented as adjoined to the verb it modifies.

In some models, such as that of Abney (1989), if an attachment ambiguity arises during parsing, it is resolved in favor of treating the ambiguous phrase as an argument. Thus, when interpreting a phrase like *interested the man in a car*, the fact that *in a car* potentially is an argument of *interest* but an adjunct of *man* should guide the parser to choose the former option. In another approach, Frazier and Clifton (1996) propose a theory of parsing in which "primary phrases" (which include arguments; more generally, primary phrases are the subject and main predicate of a finite clause and any complements or obligatory dependents of a primary phrase) and "non-primary phrases" (including adjuncts) are parsed in quite distinct ways. The parser is claimed to attach a potential primary phrase into a determinate position in a phrase marker, following structural parsing principles described by Frazier (1979; cf. Frazier, 1987, 1990; Frazier & Rayner, 1982). How-

This research was supported in part by NIH Grant HD-18708 to the University of Massachusetts, NIMH Grant R29 MH51768-01 to Northeastern University, and NIMH Grant R29 MH51768-02 to the University of Kansas. Portions of the results reported here were reported to the Psychonomic Society in 1991. Our thanks to Shelia Kennison, Lyn Frazier, Maryellen MacDonald, and Peter Dixon for their helpful comments on earlier versions of this paper. We thank the HCRC, University of Edinburgh, for their support in providing the second author with an environment in which this report could be completed. Correspondence should be addressed to S. R. Speer, Department of Speech, Language, and Hearing Sciences, Dole Center, University of Kansas, Lawrence, KS 66045 (e-mail: speer@ukans.edu).

—Accepted by previous editor, Geoffrey R. Loftus

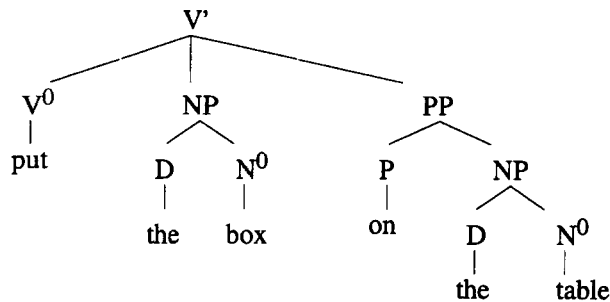
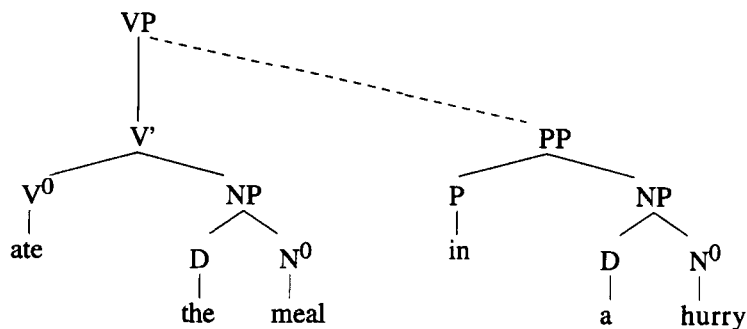
PP argument of VPP adjunct of V

Figure 1. Phrase structure diagrams for a PP argument (top panel) and a PP adjunct of a verb (bottom panel). The dashed line indicates association of the PP with the domain of the VP.

ever, it simply “associates” a nonprimary phrase in a syntactically underspecified fashion with a larger domain of a sentence, within which it can be interpreted. This association is depicted in Figure 1 by a dashed line.

Several sources of evidence suggest that the argument/adjunct distinction and related distinctions do play a role in parsing. For example, Britt (1994) showed that discourse context affects the parsing of obligatory and optional arguments differently. Referentially supportive contexts overrode a syntactic preference for a prepositional phrase to serve as an argument of a verb (rather than an adjunct of a noun phrase) when the argument was optional but not when it was obligatory. Most closely relevant to “prefer argument” strategies such as Abney’s (1989) is that of Clifton, Speer, and Abney (1991). These authors examined the reading of sentences with PPs which could be either argument or adjunct of either a verb or a noun. For instance, *in a wallet* is an argument of a verb in *She interested the man in a wallet* but an argument of a noun in *She expressed her interest in a wallet*. According to the “prefer argument” principle, the preferred interpretation of *in a wallet* would be to modify the verb in the first case but the noun in the second case. However, Clifton et al. found faster first-pass reading time (as described in Rayner,

Sereno, Morris, Schmauder, & Clifton, 1989) when a PP modified a verb than when it modified a noun, whether it was argument or adjunct; thus, *in a hurry* was read quickly and easily in *He expressed his interest in a hurry* as well as in the first example above, and both were read more quickly than any case in which the PP modified the noun. Reading time later in the sentence was generally faster when the PP was an argument of either noun or verb than when it was an adjunct. Thus, while there seems to be an initial parsing preference in which PPs are taken to modify verbs, there is a later, very strong, and intuitively available preference to take a PP as an argument rather than an adjunct.¹

In the present paper, we explore the advantage of verb arguments over verb adjuncts (e.g., in materials like those used by Clifton et al., 1991, the advantage of *She interested the man in a wallet* over *He expressed his interest in a hurry*). This difference may reflect the time taken by a subtle kind of reanalysis, from analyzing the PP as an argument of the verb to reanalyzing it as an adjunct of a verb. This advantage would be a point of substantial theoretical interest, consistent either with some sort of “prefer argument” principle or with a structural parsing preference that indirectly favors arguments over adjuncts (as

in Frazier & Clifton's, 1996, construal hypothesis; see the Discussion section for Experiment 1 below). However, the advantage of arguments over adjuncts in Clifton et al. (1991) may actually reflect a consequence of the experimental design that they used. In their materials, a verb adjunct PP was always preceded by a noun that could assign an argument (e.g., the noun *interest*), but a verb argument PP was always preceded by a noun that did not assign an argument (e.g., the noun *man*). Readers may have been slowed because they had to evaluate the PP as a possible argument of the noun in the former case but not in the latter case. More generally, the observed advantage of arguments over adjuncts may only reflect the greater plausibility of arguments, a factor which was not controlled in the Clifton et al. experiments.

In the experiments reported here, we examined both of these possibilities by controlling and varying plausibility and by eliminating the presence of argument-assigning nouns. Plausibility has long been recognized as an important influence on reading ease (Bever, 1970; Clark & Clark, 1977; Just & Carpenter, 1980). Certainly, since Marslen-Wilson's early work (Marslen-Wilson, 1975; Marslen-Wilson & Tyler, 1980), it has been clear that readers and listeners construct and utilize the meanings of sentences very rapidly, essentially on a word-by-word basis. Still, precisely how this is done is only poorly understood. Logically, it must be the case that the plausibility of an analysis of a sentence, not of a string of words, is evaluated. A given NP, for example, is plausible or implausible as (say) a subject or an agent of a verb; its plausibility cannot be determined without specifying its grammatical relation to the verb. Therefore, one or more grammatical analyses ("parses") of the relevant portion of a sentence must be available before plausibility can have an effect on reading. We suggest that determining whether a phrase is an argument or an adjunct is a necessary preliminary to evaluating its plausibility (specifically, its plausibility as an argument or as an adjunct of some particular word or phrase).

There is debate, however, about just how the status of a phrase as argument versus adjunct affects parsing, and it is even possible to question whether argument status and rated plausibility are in fact distinct. Models such as that of MacDonald, Pearlmutter, & Seidenberg (1994a) do not make an explicit distinction between arguments and adjuncts and simply claim that arguments are more strongly represented than adjuncts in lexical entries. We assume that more strongly represented phrases will generally be judged to be more plausible. In a similar vein, Spivey-Knowlton and Sedivy (1995) have suggested that the difference which Britt (1994) observed between optional and obligatory arguments may simply have reflected differences in the strength of a "frequency-based or a semantically-based lexically specific constraint" (p. 256) of the sort that they argued accounted for their own findings about the comprehension of different verb phrase modifiers. Although Spivey-Knowlton and Sedivy focus their discussion on the resolution of competition between

alternative analyses with different strengths, a reasonable extension of their claims would hold that what we call adjuncts, optional arguments, and obligatory arguments differ simply in the strength of the constraints supporting them and hence should differ in rated plausibility or acceptability.

The present experiments were designed to search for distinct effects of argument status and plausibility. We matched arguments and adjuncts on plausibility, varying plausibility factorially with argument status to see where in parsing each would have its effect. We compared PPs that were arguments and adjuncts of a verb under experimental conditions in which no competing argument-assigner was present. Finally, we used both self-paced reading and eyetracking methodologies to provide converging evidence and to attempt to pinpoint just where in processing the variables of argument status and plausibility would have their effects.

Thirty-four quadruples of items like those shown in Table 1 were constructed. Each of the sentences contained a subject, a transitive verb, and a direct object NP that could not accept a PP as an argument. In the third example, the NP *the toxic waste dump* is not an argument assigner. Two versions of each item had PP verb arguments, and two had PP verb adjuncts. All four phrases used with a given item had the same preposition.

Our definition of argument versus adjunct was much the same as that in Clifton et al. (1991). An argument phrase had to be lexically dependent on the verb, at least in the sense that its interpretation depended on the verb (but the argument was not an obligatory argument in any case). An adjunct phrase, on the other hand, was not lexically dependent on the verb. Adjuncts, for the most part, semantically expressed accompaniment, manner, location, justification, number, and time or duration. The semantics expressed by arguments, of course, differed among the verbs used. After we constructed our items, Schütze's (1995; Schütze & Gibson, 1996) recommendations for argumenthood diagnostics came to our attention. We applied four of these diagnostic tests, designed to distinguish arguments from adjuncts, to the 20 sets of items we finally selected according to the procedures described below. The tests we used were "order" (arguments appear before adjuncts in a fully acceptable string), "iterativity" (one cannot iterate arguments but one can iterate adjuncts; cf. **I rented the flat to yuppies, to libertarians* but *I met a student with blue eyes, with a wonderful smile*), "pro-form replacement" (in replacing a head with a pro-form, e.g., *do so*, one must include an argument but need not include an adjunct; cf. **John described the film to Mary, and Fred did so to Sue* but *John filled out the form in pen, and Mary did so in pencil*), and "separation from the head" (a preposed PP cannot be followed by a question if it is an argument, but can if it is a modifier; cf. **On the shelf, who put the book?* but *On Tuesday, who drove to the store?*). We did not use Schütze's wh-extraction test because it did not give clear-cut results on our materials, or on others that we tried. Each of our pairs of argument items passed at least

Table 1
Sample Sentences and Mean Sensibleness Ratings
(SDs in Parentheses)

Argument, High Plausible:	1.56 (0.374)
The people who lived near Love Canal blamed the toxic waste dump for <i>their leukemia</i> , but they never had enough resources to sue.	
The wealthy investor paid ten thousand dollars for <i>a ski vacation</i> , and never missed the money.	
Argument, Low Plausible:	3.67 (0.756)
The people who lived near Love Canal blamed the toxic waste dump for <i>their hairdos</i> , but they never had enough resources to sue.	
The wealthy investor paid ten thousand dollars for <i>free samples</i> , and never missed the money.	
Adjunct, High Plausible:	2.14 (0.338)
The people who lived near Love Canal blamed the toxic waste dump for <i>several years</i> , but they never had enough resources to sue.	
The wealthy investor paid ten thousand dollars for <i>reasons of conscience</i> , and never missed the money.	
Adjunct, Low Plausible:	3.49 (0.380)
The people who lived near Love Canal blamed the toxic waste dump for <i>a few moments</i> , but they never had enough resources to sue.	
The wealthy investor paid ten thousand dollars for <i>the heck of it</i> , and never missed the money.	

Note—Argument/adjunct phrases are in italics. Rating scale: 1 = most sensible, 5 = nonsensical.

three of the four tests for argumenthood, and each of our pairs of adjunct items passed at least three of the four tests for adjuncthood. All 20 of our argument items passed the iterativity and head separation tests for arguments, and all 20 of our adjunct items passed the order test for adjuncts (and 18 of 20 passed the pronoun replacement test). When an item did not pass a test, it did not satisfy the requirements of the other type of item. Rather, the outcome of the test (in the judgment of one or both of the authors) was simply too unclear to conclude whether it passed or failed. The most problematic test (apart from the wh-extraction test that was not used) was the head separation test, which resulted in 9 of 20 “uncertain” decisions for adjunct items). Each item had a plausible and an implausible version of both the argument and adjunct phrases. Sentences were plausible or implausible as whole sentences, but no sentence was semantically anomalous.

We pretested high- and low-plausibility argument and adjunct versions of these 34 items in a plausibility rating task modeled after one used by Taraban and McClelland (1988). Our pretest subjects were given the beginning of

an item to complete (e.g., *The people who lived near Love Canal blamed the toxic waste dump for . . .*). Immediately after writing their completion, they were shown one of our own (argument vs. adjunct × high vs. low plausibility). They then rated the completion for sensibleness on a five-point scale, where 1 was *the experimenter’s completion is as sensible as mine* and 5 was *the experimenter’s completion is nonsensical*. Four groups of 17 subjects each (undergraduates at Northeastern University and at the University of Massachusetts, Amherst) were used. Each group rated one quarter of the 136 completions, in a counterbalanced fashion, so that each subject rated only one completion of each item and each subject rated approximately equal numbers of each type of completion.

We chose the 20 items that best matched arguments and adjuncts in rated sensibleness and provided a big difference between high- and low-plausible items. Items and their ratings are listed in Appendix A. The mean sensibleness ratings of these 20 items (1 = most sensible, 5 = nonsensical) appear in Table 1. The means show that the plausible conditions were rated as far more sensible than the implausible conditions. While there was no difference in ratings for the implausible argument versus implausible adjunct sentences, plausible argument sentences were somewhat more highly rated than plausible adjuncts. The former mean difference was 0.01, and the latter was 0.60.

The experimental materials differed slightly among conditions in mean length of the critical PP region, and more substantially in lemma frequency (Francis & Kučera, 1982). The values appear in Table 2. The differences in length are small enough to be dismissed in interpreting any effects observed. The differences in lemma frequency do mirror differences among the experimental conditions and may qualify the results. To anticipate the results, lemma frequency of the main noun of the object of the preposition was not significantly correlated with reading time. Because of this, we will argue that the results (or the lack of results) observed cannot be attributed to differences in length or frequency.

EXPERIMENT 1

In Experiment 1, a self-paced reading procedure was used to measure the time taken to read sentences like those displayed in Table 1. Highly plausible sentences should, of course, be read more quickly than less plausible sentences. Our main focus was on the effect of argument

Table 2
Mean Length (With SDs; Characters Including Spaces) of PPs
and Mean Log Lemma Frequency (Francis & Kučera, 1982)
of the Main Noun of the PP Object NP, Experiments 1–2

Measure	Condition							
	Arg, Hi Plaus		Arg, Lo Plaus		Adj, Hi Plaus		Adj, Lo Plaus	
	M	SD	M	SD	M	SD	M	SD
Mean length	18.7	4.05	18.5	4.40	18.55	4.51	18.1	3.88
Mean log frequency	1.70	0.83	1.21	0.76	2.09	0.618	1.46	0.80

Note—Arg, argument; Adj, adjunct; Hi Plaus, high plausibility; Lo Plaus, low plausibility.

Table 3
Reading Times (in Milliseconds and in Milliseconds/Character), Experiment 1

Condition	Region									
	1		2		3		4		5	
	(Subject NP)		(Verb NP)		(Critical PP)		(Next)		(Next + 1)	
	Msec	Msec/Char	Msec	Msec/Char	Msec	Msec/Char	Msec	Msec/Char	Msec	Msec/Char
Arg, Hi Plaus	762	65.0	925	51.5	791	49.0	788	50.3	847	61.0
Arg, Lo Plaus	761	65.0	888	49.8	874	54.1	901	56.8	883	63.0
Adj, Hi Plaus	752	62.0	882	50.4	812	51.9	798	50.0	806	58.0
Adj, Lo Plaus	759	65.0	924	52.0	916	62.2	844	57.7	754	62.0

Note—Pooled *SEs* are estimated from analyses of variance error term for Region 3 of the milliseconds measure: 21.4 (subjects) and 25.1 (items). The values were 1.089 and 1.245, respectively, for the milliseconds/characters measure. Arg, argument; Adj, adjunct; Hi Plaus, high plausibility; Lo Plaus, low plausibility.

versus adjunct status and its interaction with plausibility. We predicted that if the parser preferred argument over adjunct structure for verbs, reading times should be faster in the argument conditions than in the adjunct conditions for both plausible and implausible sentences. On the other hand, if our previous finding of faster reading for verb arguments than for verb adjuncts (Clifton et al., 1991) was due simply to general plausibility or to the extent to which overall "content" of the sentences supported interpretation of the PP, reading times should be faster for plausible sentences than for implausible sentences, regardless of the argument/adjunct distinction. Finally, we raised the possibility that our previous finding of an argument preference might have been due to inflated times in the adjunct condition caused by interference from the argument-assigning noun in direct object position. If this were the case, we should find no difference between arguments and adjuncts in the sentences used in the present study, because they did not contain such nouns.

Method

Materials. The 20 quadruples of sentences chosen in the pretest were used as experimental items. Two examples of the four forms of a sentence appear in Table 1, and all items appear in Appendix A. Ninety-six other sentences were included in the list to be presented in the experiment. Twenty of these items came from other, unrelated experiments (one on ambiguity of attachment of a PP to one of two nouns, and the other on a direct object/sentence complement ambiguity), and 76 were filler sentences of a wide variety of syntactic constructions. Half of the experimental items, and 78 of the remaining items, were followed by a simple true-false question. Only a few of the questions of experimental items questioned the critical PP. A practice list of eight items, composed of materials similar to those used in the main experiment, was also constructed.

Procedure. The subjects first received the practice list to familiarize themselves with the procedures. Then the 116 items in the experimental list were presented to them in an individually randomized order. A microcomputer presented the sentences one phrase at a time in a self-paced reading task, in "moving windows" with a noncumulative display (Kennedy & Murray, 1984). The subjects pressed a thumb button to advance the sentence from one presentation region to the next. A question, if scheduled, was presented immediately after the last region of each sentence, all at once, with the options *true* and *false* displayed on the video screen. The subjects indicated which answer was correct by pulling one of two triggers.

Subjects. Eighty undergraduates at the University of Massachusetts, all native speakers of English, were tested in individual 40-min sessions. They received course credit for their participation.

Results

Accuracy on the true-false questions averaged 83% for adjuncts and 87% for arguments, with no difference between plausible and implausible items (84% vs. 85%, respectively). Reading times are shown in Table 3, and the effects of interest are summarized in Figure 2 (which contains error bars indicating ± 1 *SE*, calculated using a method recommended by Bakeman & McArthur, 1996, to estimate the variability in RTs within each individual condition after removing the overall differences in mean reading time between subjects). The reading times are reported both in milliseconds/character and in milliseconds in Table 3. The phrases being compared were closely matched in length (see Table 2), so that no correction is needed for differences in length (cf. Clifton & Ferreira, 1987, and Trueswell, Tanenhaus, & Kello, 1993, for discussion of some of the problems of using milliseconds/character as such a correction). However, the milliseconds/character measure does reduce error variance by making different items roughly comparable with one another. Therefore, statistical tests were performed on the milliseconds/character data.

Reading times for the PP were faster when the PP was an argument than when it was an adjunct [51.5 vs. 56.6 msec/character, $F_1(1,79) = 37.61, p < .02$; $F_2(1,19) = 7.65, p < .02$]. They were also faster when the PP was plausible than when it was implausible [50.2 vs. 57.9 msec/character, $F_1(1,79) = 42.69, p < .001$; $F_2(1,19) = 13.81, p < .002$]. The interaction between plausibility and argument status was significant by subjects and nearly significant by items [$F_1(1,79) = 5.68, p < .02$; $F_2(1,19) = 3.71, p < .08$]. The argument effect was significant, or nearly so, for each level of plausibility [$t_1(79) = 5.21, p > .01$; $t_2(19) = 2.76, p < .02$, for low plausibility; $t_1(79) = 2.31, p < .03$; $t_2(19) = 1.71, p < .10$, for high plausibility]. In the region after the PP, the plausibility effect remained large and significant [$F_1(1,79) = 53.18, p < .001$; $F_2(1,19) = 20.29, p < .01$], while the argument advan-

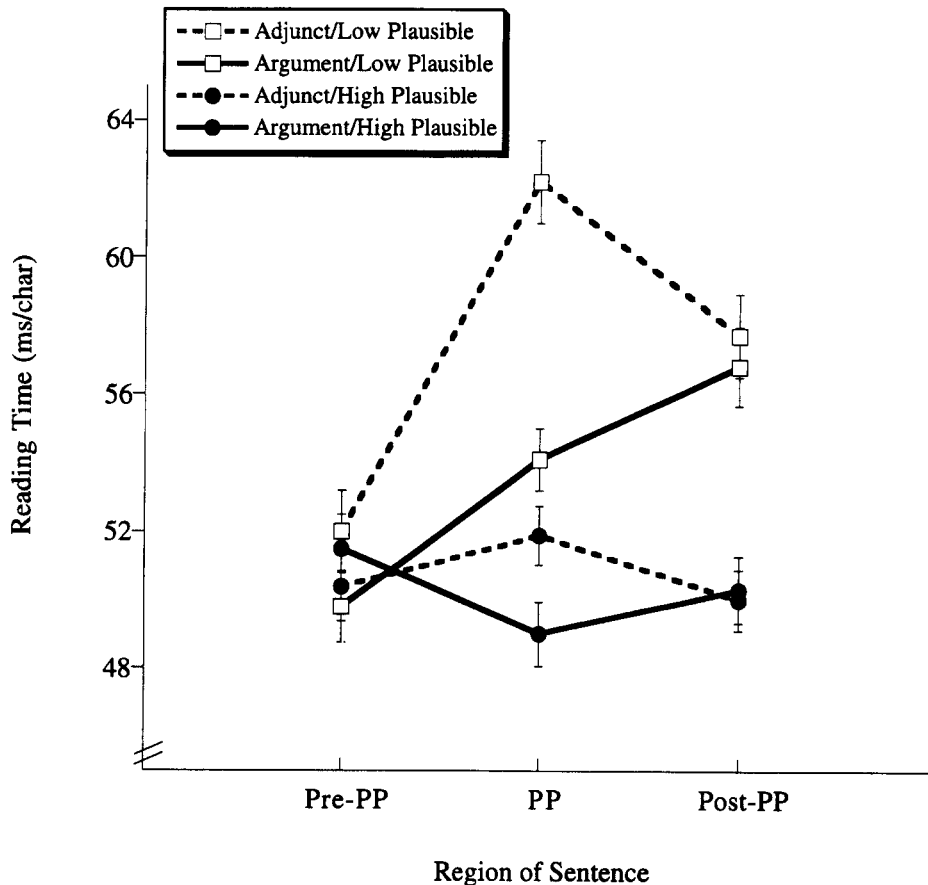


Figure 2. Reading times (milliseconds/character) in each of three regions of the sentences, Experiment 1. *SEs* (estimated from the pooled error terms of an analysis of variance treating argument status, plausibility, and region as factors) were 1.08 msec/character (subjects) and 1.51 msec/character (items).

tage disappeared for both high- and low-plausible items ($F < 1$).

However, the argument advantage for high-plausible items is confounded with a difference in plausibility: High-plausible arguments were rated as more plausible than high-plausible adjuncts. The existence of a clear effect of argument versus adjunct for low-plausible items, where there was no confounding plausibility difference, indicates that the argument advantage cannot be accounted for entirely as an effect of plausibility. This conclusion is buttressed by the results of an analysis of covariance in which rated plausibility was the covariate in an analysis permitting generalization to items (treating both argument vs. adjunct and plausible vs. implausible as between-items variables, since different sentences with different plausibility ratings were used in each cell of the design). The advantage of arguments over adjuncts remained significant in this analysis [$F_2(1,75) = 5.37$; $SE = 1.80$], and the interaction between argument versus adjunct and plausible versus implausible became significant [$F_2(1,75) = 4.30$]. The mean reading times (in milliseconds/character), adjusted for the covariate of rated plausibility, appear in

Appendix B. Note that the argument effect nearly disappeared for high-plausible items, although it remained significant overall. We must conclude that the argument effect that was apparently seen for highly plausible items in the initial analysis may have reflected the confound with plausibility, but we can also conclude that there is a true argument advantage at least for relatively implausible items.

As is shown in Table 2, the four conditions differed substantially in the mean Francis-Kučera lemma frequency of the head noun of the NP object of the preposition. Higher frequency nouns might be expected to result in faster reading. If this were the case, the reading time difference between high and low plausibility might be artifactually inflated, and the reading time difference between arguments and adjuncts, artifactually attenuated. However, an analysis of covariance of mean reading times (milliseconds/character, critical PP), with log noun frequency as the covariate (treating conditions as a between-items factor because different nouns appeared in the different conditions) continued to indicate a significant effect of plausibility [$F_2(1,72) = 10.80$, $p < .002$] as well as argument status [$F_2(1,72) = 9.23$, $p < .01$]. (The *df* reflect

the fact that there were three missing cases, for words that did not appear in the frequency tables.) In fact, the correlations between log lemma frequencies and reading times, calculated separately for each experimental condition, averaged a very small $-.14$, presumably because other factors affecting reading time were not controlled across the nonsystematic variability in frequency.

Discussion

Why was there an advantage of arguments over adjuncts? We propose that the advantage has structural causes. It is not a reflection of simple differences in plausibility between arguments and adjuncts (although it might have been in Clifton et al., 1991). Although we did observe a large effect of plausibility, the argument advantage could not be reduced to a plausibility effect. Furthermore, the argument advantage is not an artifact of the presence of a competing argument assigner in the case of verb–adjunct PPs, as it might have been in Clifton et al. (1991). The present sentences contained no such competing argument assigner, but nonetheless elicited faster reading times when the PP was an argument of the verb than when it was an adjunct.

The argument advantage does not fall immediately out of principles that we have defended previously. In particular, *minimal attachment* (Frazier, 1979, 1987) by itself does not result in a preference for verb arguments rather than adjuncts. As Figure 1 shows, a PP can in principle be attached into a VP either as an argument (sister of V, top panel) or an adjunct (sister of V', bottom panel; we are assuming that multiple branching structures are possible). Either attachment is minimal, in that each requires the addition of the same number of syntactic nodes. We would like to suggest several possible explanations for the argument advantage. The first is an elaboration of the *late closure principle* (Frazier, 1979), which claims that (grammatical considerations and the minimal attachment principle permitting), a new word is attached into the phrase currently being processed. Attachment of the PP into the NP that is being processed is blocked because it is nonminimal. Dismissing this NP, the currently processed phase is the "V-bar" phrase, V' in Figure 1. Attaching the PP into this phrase as a sister of V would make the PP an argument of the V. If the PP were inappropriate as an argument, revising it to be an adjunct might take time.

Another possibility is that a "prefer argument" principle is correct. Clifton et al. (1991) argued against this possibility, because of their finding of an initial preference for verb adjuncts over noun arguments. However, they acknowledged that their finding could be ascribed to the existence of some other factor favoring verb modification in English, which could compete with an underlying general preference for arguments over adjuncts.

A third, and novel, possibility is suggested by a different way of looking at the interaction of reading times observed at the PP. The plausibility effect was larger for the adjunct than for the argument sentences, even though the difference in rated plausibility was smaller. Adjunct

phrases may rely heavily on factors of world knowledge for their interpretation, whereas argument phrases rely more on grammatical information. The latter type of information may simply be processed faster and in a more homogeneous fashion, resulting in a brief argument advantage and a decrease in the size of the plausibility effect for arguments as compared with adjuncts. We cannot distinguish among these three possibilities at the present time. However, we believe that we can reject a fourth possibility. This is the possibility that any PP is initially attached in an unspecified manner within the VP, in effect spreading over all possible attachment sites so that its interpretation is made precise, if necessary, on the basis of plausibility considerations. Such a possibility (which is essentially the *pseudoattachment* proposal of Church, 1980) would provide no particular advantage for arguments over adjuncts, no particular advantage for verb attachments over noun attachments, and certainly no advantage that is independent of that for plausible over implausible attachments.

Rejecting this possibility is consistent with the *construal hypothesis* being developed by Frazier and Clifton (1996; cf. Carreiras & Clifton, 1993; Gilboy, Sopena, Clifton, & Frazier, 1995). The construal hypothesis distinguishes between *primary* and *nonprimary* phrases. A primary phrase is the subject or predicate of a finite clause, or any complement or obligatory dependent of a primary phrase. As stated earlier, any phrase that could potentially be a primary phrase is attached in a determinate fashion into a phrase structure tree, whereas a phrase that must be a nonprimary phrase is "associated" in an underspecified fashion with the tree, permitting nonsyntactic and nongrammatical sources of information to determine its interpretation. Since a PP is a potential primary phrase, it will initially be attached into the tree, which (following late closure, as outlined above), will result in its being analyzed as an argument (e.g., top panel, Figure 1). Only if later interpretation disconfirms this analysis will it be treated as a nonprimary phrase, a phrase that is loosely associated with a domain such as a VP and eventually analyzed as an adjunct (bottom panel, Figure 1).

In addition to interpretive questions, questions about the detailed nature of the data remain. The argument advantage appeared on the PP in the present study, whereas in the self-paced reading study reported by Clifton et al. (1991) it appeared only in the next region. There were many differences between the two studies, including the fact that different verbs were used in the contrast made in the earlier study, whereas the same verb was used in both conditions in the present study. We designed Experiment 2 to explore these differences further, by measuring eye movements during reading to pin down the point at which effects of argument status and plausibility appear.

EXPERIMENT 2

The second experiment was essentially the same as the first, except for the technique used to measure reading

speed. Rather than self-paced reading, we used an eyetracker to measure fixations during normal reading. The rich data provided by eye movement measurement allows a finer grained evaluation of reading effects than is generally provided by other techniques (Rayner et al., 1989).

Method

Materials and Procedure. The same 20 experimental sentences used in Experiment 1 were used in Experiment 2. The total list of sentences included 12 additional sentences with a temporary PP attachment ambiguity, 8 with a temporary direct object/sentence complement ambiguity, and 60 filler sentences of a wide variety of syntactic structures. Forty-eight of the sentences (including 10 of the experimental sentences) were followed by a simple true–false question. In addition, a practice list of 8 sentences (4 followed by questions) was constructed.

Sentences were presented on a Sony 1302 video monitor displaying 5×8 dot matrix letters generated by a Targa 16 video board. The display presented a maximum of 72 characters in each line, with approximately 4 characters per degree of visual angle. The experimental sentences were one or two lines long on the screen. The critical PP sometimes occurred on the first line and sometimes on the second, but never spanned two lines.

Subjects' reading was measured with an SRI/Fourward Technologies Generation V Eyetracker. A bite bar was prepared for each subject, and the functioning of the eyetracker was explained before the experiment began. The subjects were told that they should read in a normal fashion, attempting to understand each sentence so that they could answer straightforward questions about it, and that they should press a button when they had read it to their satisfaction. After an initial calibration period, the list of practice sentences was presented in an individually randomized order, and the subjects were instructed to read them at a comfortable rate, pressing a button after each sentence had been read. Before each sentence, a brief calibration check was performed, and the eyetracker recalibrated if necessary. During the presentation of each sentence, an 80286 micro-computer interfaced with the eyetracker sampled the eye position each millisecond, determining when and where each fixation began and ended. If a question was scheduled, it was visually presented after the sentence, and the subject answered by pressing one of two buttons for "true" or "false." After practice, the eyetracker was recalibrated, and the 96 experimental and distractor sentences were presented in individually randomized order, followed by questions when scheduled. The entire session took approximately 30–40 min.

Subjects. Thirty-six members of the University of Massachusetts community, unfamiliar with the purposes of the present experiment, were tested. Each received course credit or \$5 in cash for the single half-hour session.

Results

For purposes of the primary analyses, the sentences were divided into regions (see Appendix A). These were the same regions that had been used as presentation regions in Experiment 1 (and they were comparable to those used in Clifton et al., 1991). The regions were the subject NP, the verb plus the direct object NP, the critical PP, the short region following the critical PP, and, finally, the next short phrase, sometimes but not always including the rest of the sentence. Following Rayner et al. (1989) the data were analyzed in a variety of ways, including the following:

1. First fixation (the duration, in milliseconds, of the first fixation made in a region, provided that no previous fixations had been made after that region; cases in which

no fixation was made in the region before the eyes moved past the region were ignored).

2. Spillover (the duration of the first fixation made after leaving a region).

3. First-pass time (milliseconds/character: the total summed duration of fixations made in a region prior to leaving the region or prior to any fixations after the region, divided by the number of characters in the region; again, nonfixations were ignored).

4. Cumulative region reading time (CRRT; the summed duration of all fixations made after first fixating in a region until first fixating a point after the region in the sentence; the term is taken from Brysbaert & Mitchell, 1996; Hemforth, Konieczny, Scheepers, & Strube, 1994, call it "regression path analysis," and we have referred to it previously as "go-past time").

5. Total time (in milliseconds/character: the total summed durations of all fixations made in a region between presentation and removal of a sentence; sentences where no fixations were made in a region were ignored).

6. Percent regressions out of region (the percentage of trials on which one or more fixations were made in a region where at least one fixation in a region was immediately followed by a fixation in an earlier region).

Several of these measures—first fixation, first pass, CRRT, and total time—are presented in Table 4, together with omnibus standard errors for individual means, estimated by pooling the error terms in the relevant analyses of variance. A follow-up analysis was conducted to probe some nearly significant effects of argument status observed in several of the measures listed above. In this analysis, the first fixation made on a substantive word in the critical PP (in general, the first word past the preposition and determiner, and thus the first word at which disambiguation could begin) was identified, together with the immediately preceding fixation and the immediately following three fixations. (Note that if no fixation was made on the substantive words of a PP, the region presumed to contain the substantive words was expanded by up to three characters to the left, honoring the fact that some information can be obtained from the word immediately to the right of a fixation.) The means of these individual fixation durations appear in Figure 3, with standard errors for individual means estimated as described in Experiment 1. The significant effects ($p < .05$, unless otherwise indicated) of plausibility versus implausibility will be discussed first, followed by a discussion of the effects of argument versus adjunct status. These two variables did not interact significantly in any measure of reading performance. In general, the effects to be discussed are those observed in the region of the PP and the next region (unless otherwise noted). Region (PP vs. next) often had significant effects, but since regions differed in length and in the words that they contained, these effects are largely uninterpretable and will not be discussed. Interactions between regions and the factors of interest will be discussed, however.

Before these data are presented, let us briefly consider the relation between reading time and frequency of the

Table 4
Measures of Eye Movement, Experiment 2

Condition	Region					
	PP-1	PP	Next	PP-1	PP	Next
	First Fixation (Milliseconds)			First Pass (Milliseconds/Character)		
Arg, Hi Plaus	242	245	241	25.89	25.73	26.01
Arg, Lo Plaus	238	240	259	25.27	28.74	30.93
Adj, Hi Plaus	246	227	246	27.98	26.39	28.49
Adj, Lo Plaus	239	244	273	26.38	31.71	30.67
	<i>SE</i> (1) = 8.20			<i>SE</i> (1) = 1.20		
	<i>SE</i> (2) = 8.75			<i>SE</i> (2) = 1.38		
	CRRT (Milliseconds)			Total Time (Milliseconds/Character)		
Arg, Hi Plaus	782	573	540	31.05	27.70	29.40
Arg, Lo Plaus	773	677	662	31.02	33.39	35.90
Adj, Hi Plaus	744	608	587	31.06	29.11	31.41
Adj, Lo Plaus	795	699	801	32.86	39.51	38.24
	<i>SE</i> (1) = 36.51			<i>SE</i> (1) = 1.42		
	<i>SE</i> (2) = 44.24			<i>SE</i> (2) = 1.82		

Note—*SE*s are presented for individual means as estimated from the pooled error terms of analyses of variance treating argument status, plausibility, and region (PP and Next, or in the case of First Fixation, region Next alone) as factors. Arg, argument; Adj, adjunct; Hi Plaus, high plausibility; Lo Plaus, low plausibility.

main noun of the object of the preposition. As in Experiment 1, the Francis–Kučera (1982) lemma frequencies of this noun were correlated with measures of reading time for the entire PP, separately for each experimental condition. The mean of the four correlations for first pass reading time (milliseconds/character) was .004, and the mean for total reading time (milliseconds/character) was .038. As was the case in Experiment 1, differences in frequency of occurrence (within the constrained range used in the present experiment) had a trivial effect in comparison with the effects of other variables in the experiment.

Plausibility. The plausibility of a PP as a modifier of a verb had pervasive effects, generally beginning in the PP and continuing into the next region. First fixation durations did not differ between high- and low-plausibility items in the PP region ($p > .20$), but they were longer for low-plausibility items in the region after the PP [next; $F_1(1,35) = 6.61$; $F_2(1,19) = 5.57$]. Furthermore, the spillover analysis showed that the very first fixation after leaving the PP tended to be shorter for highly plausible than for implausible PPs, averaging 239 versus 263 msec [$F_1(1,35) = 5.28$; $F_2(1,19) = 3.42$; $p < .10$].

First-pass reading times were faster for high- than for low-plausible phrases when pooled over the PP and next regions [26.7 vs. 30.5 msec/character overall; $F_1(1,35) = 17.78$; $F_2(1,19) = 11.07$], as were total reading times [29.4 vs. 36.8 msec/character; $F_1(1,35) = 53.99$; $F_2(1,19) = 26.36$]. Cumulative region reading time was also less for high- than for low-plausible items in both the PP and the next regions [577 vs. 710 msec; $F_1(1,35) = 21.84$; $F_2(1,19) = 14.89$]. Regressions out of these two regions were less frequent for high- than for low-plausible phrases [5.8% vs. 10.9%, respectively; $F_1(1,35) = 11.37$; $F_2(1,19) =$

14.51]. All these measures show some source of difficulty in reading implausible PP verb modifiers.

Now turn to the analysis of the first few fixations in the substantive region of the PP, illustrated in Figure 3. The first fixation in this region (FF) did not show a significant effect of plausibility [$F_1(1,35) = 3.44$, $p < .07$; $F_2(1,19) = 2.23$, $p > .15$]. The immediately following fixation (FF+1), however, did [$F_1(1,35) = 4.95$, $p < .04$; $F_2(1,19) = 5.30$, $p < .04$], as did the mean of this fixation and the next fixation (FF+2) [$F_1(1,35) = 6.91$, $p < .02$; $F_2(1,19) = 8.52$, $p < .01$]. Simple contrasts indicated that the plausibility effect (pooled over FF+1 and FF+2) was significant by subjects when the PP was an argument [$t_1(35) = 2.16$; $t_2(19) = 1.49$, $p > .10$] and significant by both subjects and items when the PP was an adjunct [$t_1(35) = 3.49$; $t_2(19) = 2.40$]. No interactions approached significance.

Argument versus adjunct status. Arguments and adjuncts did not differ significantly in the duration of first fixations, either in region PP or next. The “spillover” analysis showed that the duration of the very first fixation after leaving the PP was faster when that PP was an argument than when it was an adjunct [236 vs. 261 msec; $F_1(1,35) = 10.34$; $F_2(1,19) = 5.61$]. However, the difference was significant ($p < .05$) only for low-plausible items (238 vs. 277 msec), not for high-plausible items (235 vs. 245 msec). All other effects showed a numerical advantage of arguments over adjuncts, but none of these effects was fully significant. Cumulative region reading time and total time most closely approached significance [$F_1(1,35) = 12.32$; $F_2(1,19) = 2.77$, $p < .12$; and $F_1(1,35) = 10.12$, $p < .05$; $F_2(1,19) = 3.36$, $p < .10$, respectively]. The numerical advantage was consistently present for both high- and low-plausibility items, but as was the case

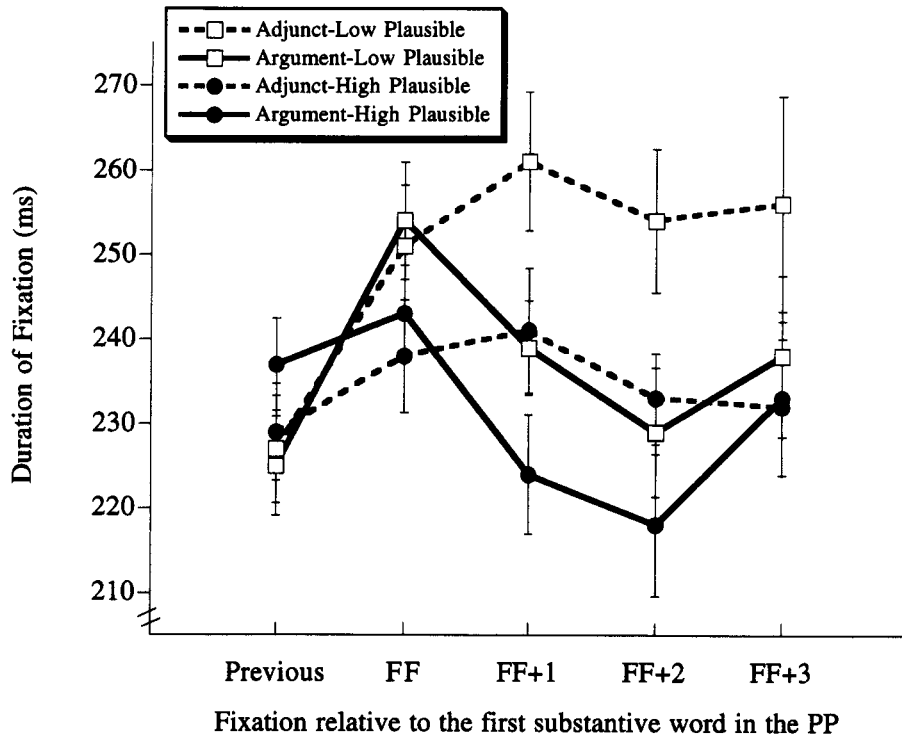


Figure 3. Reading times (milliseconds) for the first fixation (FF) on the first substantive word in the PP, the preceding fixation, and the three following fixations, Experiment 2. SEs (estimated from the error terms of an analysis of variance for the means of FF+1 and FF+2) were 7.54 msec (subjects) and 8.18 msec (items).

in Experiment 1, we can claim secure evidence for its existence only for the low-plausibility items.

The analysis of the initial fixations in the substantive region of the PP, however, provided more secure evidence for the argument status effect. While the very first fixation (FF) did not show a significant effect of argument status ($F < 1$), the immediately following fixation (FF+1) did [$F_1(1,35) = 8.85, p < .01; F_2(1,19) = 4.75, p < .05$], as did the mean of this fixation and the next one (FF+2) [$F_1(1,35) = 15.39, p < .001; F_2(1,19) = 5.90, p < .03$]. The simple contrast of argument versus adjunct was significant by subjects for high-plausibility items [$t_1(35) = 2.83, p < .01; t_2(19) = 1.94, p < .10$] and significant by both subjects and items for low-plausibility items [$t_1(35) = 3.99, p < .001; t_2(19) = 2.74, p < .02$]. An analysis of covariance of the first two fixations immediately following the substantive word (FF+1 and FF+2), using individual sentences as the random factor and treating plausibility as a covariate, also indicated that the overall effect of argument versus adjunct was significant [$F_2(1,76) = 6.72, p < .02, SE = 7.17$]. The adjusted means resulting from this analysis appear in Appendix B. Although a difference between argument and adjunct is apparent both at high and low plausibility, it is significant only in the latter case.

CONCLUSIONS

The apparent empirical conclusion from the reported experiments is that the plausibility and the argument status of a PP modifier of a verb both affect reading speed. Less plausible PPs are read substantially more slowly than more plausible ones. The effect appears during reading of the PP itself, even on first-pass reading time in the eye-tracking experiment, and persists into the following region as well. Adjunct PPs are also read more slowly than argument PPs, but this effect is smaller and shorter lived. Although it appeared in both the self-paced reading and the eye-tracking experiments, it was largely limited in the latter experiment to a few fixations following the initial fixation in the disambiguating region. It is a quick effect, but not a large one.

An interesting interaction between plausibility and argument status appeared in Experiment 1. The effect of plausibility was greater for adjuncts than for arguments (in spite of the fact that the difference in plausibility ratings was somewhat greater for arguments than for adjuncts). If the interaction could be trusted, it might mean that the interpretation of arguments is more dependent on grammatical (at least, lexically specified) knowledge, whereas the interpretation of adjuncts is more dependent

on world knowledge, which is presumably indexed by the plausibility ratings. However, the interaction failed to reach significance in Experiment 2. Some evidence of it can be discerned in some measures, such as second pass times and total times, but even when it approached significance in a by-subjects analysis [$F_1(1,35) = 3.54, p < .07$, second pass times] it did not approach significance in the corresponding by-items analysis ($p > .20$). This lack of significance makes further speculation about the basis of the interaction unwise.

One conclusion is clear. As Marslen-Wilson (1975 and elsewhere) claimed, semantic interpretation happens very quickly indeed. In our experiments, the effects of plausibility were measurable during the reading of the phrase whose plausibility was manipulated; they were not delayed. Just how early such effects on processing may be detected seems to depend on a variety of factors, not the least of which is the strength of the plausibility manipulation. Although some recent experiments examining the influence of plausibility on the processing of primary phrases have demonstrated first-pass plausibility effects (Traxler & Pickering, 1995), experiments manipulating a narrower range of plausibility have shown early effects for only a subset of syntactic conditions (Garney, Pearlmuter, Myers, & Latocky, 1997), and some more subtle plausibility manipulations have produced effects detectable only in total reading time measures (Schmauder & Egan, 1995, 1996). The plausibility manipulation in the present experiments was quite strong, and in contrast to the sentences in the previously mentioned studies, many of our sentences remained implausible even when they were complete. Thus it is perhaps not surprising to find such early and pervasive effects of semantic interpretation. However, this conclusion merely sets the stage for interesting discussion. How does the language processor integrate grammatical knowledge, lexical knowledge, and world knowledge so that it can quickly identify a grammatically licensed interpretation and evaluate it against general knowledge?

We will propose the following account of the processing of our materials, largely as a heuristic to guide further research. The account is based on Frazier's (1979, 1987) proposal of a serial parser, and on its elaboration as the construal hypothesis (Frazier & Clifton, 1996). When a postverbal PP is read, it is treated as a potential argument or primary phrase, and it is attached in a determinate fashion as the argument of the preceding verb. Interpretation begins essentially as soon as there is a structure to interpret. The plausibility of the PP as an argument is evaluated by comparing world knowledge about the contents of the postverbal PP with the lexical requirements of the verb. If the PP is an implausible argument of the verb, the parser continues to evaluate it and to explore alternative analyses.

The PP will be an implausible argument for the low-plausibility argument cases and for adjuncts in general. The evaluation of alternative analyses that takes place in these cases will presumably slow reading time. One al-

ternative that is presumably tried treats the PP as an adjunct of the verb phrase, not an argument of the verb. This will yield a plausible interpretation in the case of the high-plausible adjuncts, ending the exploration of alternative analyses. However, the parser may for some period of time continue to seek alternative interpretations for low-plausible argument and adjunct items. Reading may be slowed in these cases, especially in the case of low-plausible adjuncts, since the verb provides no lexically specific constraints on the interpretation of such items.

The crucial claims in the process just described are the following: (1) Syntactically structured phrases, not collections of words, are evaluated for their plausibility; (2) a single structural analysis is evaluated at a time; (3) the initially preferred structure treats the PP as an argument or a primary phrase; and (4) reading of a phrase for which no satisfactory interpretation has yet been found is slowed. This combination of claims predicts disruption of reading for adjuncts as compared with arguments, and for implausible items as compared with plausible ones. It does not make strong claims about the time course of this disruption, but the data that have been presented suggest that it begins in all cases shortly after the PP is read and continues for the next few fixations. It does, however, predict that arguments should be faster than adjuncts at both levels of plausibility. The data were merely consistent with this prediction in Experiment 1 (the only convincing effect being observed for low-plausible items), but the analysis of the initial few fixations in the disambiguating region of the PP in Experiment 2 provided solid evidence that arguments were faster than adjuncts at both levels of plausibility.

Other models are no doubt possible, including models cast within the *constraint-satisfaction* framework (MacDonald et al., 1994a; MacDonald, Pearlmuter, & Seidenberg, 1994b; McClelland, St. John, & Taraban, 1989; Trueswell & Tanenhaus, 1994). Such models will contrast with the process that we have described, in several ways. In their simplest form, they will evaluate multiple alternatives in parallel against multiple sources of information, including information about lexical argument structure and world knowledge. Presumably, they will predict reading time to be the slowest when there is maximal conflict among the different sources of information or perhaps when there is maximal competition among distinct analyses. Any particular analysis will be favored if it is highly plausible and if it has specific lexical support (in our terms, if it is an argument).

We find such models intriguing, but, as they are specified at the present time, unsatisfying. Some constraint-satisfaction models (see, e.g., McClelland et al., 1989) treat all sources of information alike, failing to provide theoretical devices that could support a more structured use of different types of information. Such models do not honor our claim that a grammatical analysis of a string of words is always what receives semantic interpretation, not the string of words itself. They seem capable of creating semantic relations between words that are unrelated

in a sentence's syntax as well as between words that are in grammatical construction with one another. However, we acknowledge that this shortcoming is not necessarily true of constraint-satisfaction models.

Another concern is directed at current and more sophisticated constraint-satisfaction models. This concern stems from our belief that some linguistic knowledge is stored as general rules, while other linguistic knowledge is stored as lexically specific information. We claim that these two types of knowledge can play different roles in parsing. In the present instance, we propose that general linguistic knowledge is used to construct the initial analysis of a postverbal PP as the argument of a verb, and if necessary to reanalyze it as an adjunct. Lexically specific knowledge about the argument structure of a verb is used to evaluate the initial analysis. The distinctions that we draw between lexically specific and general linguistic information, on the one hand, and between argument and adjunct relations, on the other, stand in contrast to the claims of current constraint-satisfaction theories (specifically, MacDonald et al., 1994a, 1994b). MacDonald et al.'s theory claims that all grammatical information is placed in the lexicon. Grammatical structures are always built by activating structures stored in the heads of phrases. This means that there is no essential difference between adjuncts and arguments. All information about the structure of adjunct phrases must be stored as a separate weighted link or pointer from each and every lexical entry to each and every category of adjunct phrase (or, even less economically, to each and every individual adjunct phrase). Such massively redundant representation of grammatical knowledge is theoretically inelegant. Worse, it seems likely not to provide a satisfying account of how we are able to deal productively with novel words. Connections must be formed between a newly learned verb and each of the possible classes of adjunct phrases with which it can occur. However, it seems that as soon as the verb (e.g., *to fax*) is added to our lexicon, we are able to say and understand *fax it tomorrow*. (Of course, a constraint satisfaction theorist could reply that such broadly applicable knowledge is actually associated with the category "V[erb]," and that the lexical entry for a particular verb simply makes this category available. This claim seems to accept our position that adjuncts can apply generally to verbs and to deny that all grammatical knowledge is stored with individual lexical entries.) In addition, the basic lexicalist claim seems to reduce the adjunct-argument distinction to a difference in the strength with which the relevant information is activated by a word or string of words. On the face of it, this seems likely to present some difficulty in accounting for the different effects of the argument/adjunct distinction on reading time and on plausibility ratings; why would a difference in strength of activation that affected reading time not appear in plausibility ratings?

We have no doubt that some constraint-satisfaction account of our data can be constructed. We acknowledge that the creation of more adequate constraint-satisfaction models is a useful enterprise, and we suggest that ac-

counting for data like those which we have presented might be a useful challenge to constraint-satisfaction theorists. However, we also note that the process that we have described of interpreting and evaluating syntactic structures that are constructed serially in a rule-governed manner gives a straightforward account of our data.

REFERENCES

- ABNEY, S. (1989). A computational model of human parsing. *Journal of Psycholinguistic Research*, **18**, 129-144.
- BAKEMAN, R., & MCARTHUR, D. (1996). Picturing repeated measures: Comments on Loftus, Morrison, and others. *Behavior Research Methods, Instruments, & Computers*, **28**, 584-589.
- BEVER, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279-352). New York: Wiley.
- BRITT, M. A. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory & Language*, **33**, 251-283.
- BRYNSBAERT, M., & MITCHELL, D. C. (1996). Modifier attachment in sentence processing: Evidence from Dutch. *Quarterly Journal of Experimental Psychology*, **49A**, 664-695.
- CARRERAS, M., & CLIFTON, C., JR. (1993). Relative clause interpretation preferences in Spanish and English. *Language & Speech*, **36**, 353-372.
- CHURCH, K. (1980). *On memory limitations in natural language processing*. Unpublished master's thesis, MIT.
- CLARK, H. H., & CLARK, E. V. (1977). *Psychology and language*. New York: Harcourt Brace Jovanovich.
- CLIFTON, C., JR., & FERREIRA, F. (1987). Discourse structure and anaphora: Some experimental results. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 559-586). Hillsdale, NJ: Erlbaum.
- CLIFTON, C., JR., SPEER, S. R., & ABNEY, S. (1991). Parsing arguments: Phrase structure and argument structure as determinants of initial parsing decisions. *Journal of Memory & Language*, **30**, 251-271.
- CROCKER, M. (1994). On the nature of the principle-based sentence processor. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 245-266). Hillsdale, NJ: Erlbaum.
- FRANCIS, W. N., & KUČERA, H. (1982). *Frequency analysis of English usage*. Boston: Houghton-Mifflin.
- FRAZIER, L. (1979). *On comprehending sentences: Syntactic parsing strategies*. Bloomington, IN: Indiana University Linguistics Club.
- FRAZIER, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 559-586). Hillsdale, NJ: Erlbaum.
- FRAZIER, L. (1990). Exploring the architecture of the language system. In G. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 409-433). Cambridge, MA: MIT Press.
- FRAZIER, L., & CLIFTON, C., JR. (1996). *Construal*. Cambridge, MA: MIT Press.
- FRAZIER, L., & RAYNER, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, **14**, 178-210.
- GARNSEY, S. M., PEARLMUTTER, N. J., MYERS, E., & LOTOCKY, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory & Language*, **37**, 58-93.
- GILBOY, E., SOPENA, J., CLIFTON, C., JR., & FRAZIER, L. (1995). Argument structure and association preferences in Spanish and English compound NPs. *Cognition*, **54**, 131-167.
- HEMFORTH, B., KONIECZNY, L., SCHEEPERS, C., & STRUBE, G. (Eds.) (1994). *First analysis, reanalysis, and repair*. Freiburg: Albert-Ludwigs-Universität Freiburg, Institut für Informatik und Gesellschaft.
- JUST, M. A., & CARPENTER, P. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, **85**, 109-130.
- KENNEDY, A., & MURRAY, W. S. (1984). Inspection times for words in syntactically ambiguous sentences under three presentation condi-

- tions. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 833-847.
- MACDONALD, M. C., PEARLMUTTER, N. J., & SEIDENBERG, M. S. (1994a). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, **101**, 676-703.
- MACDONALD, M. C., PEARLMUTTER, N. J., & SEIDENBERG, M. S. (1994b). Syntactic ambiguity resolution as lexical ambiguity resolution. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 123-154). Hillsdale, NJ: Erlbaum.
- MARSLÉN-WILSON, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, **189**, 226-228.
- MARSLÉN-WILSON, W. D., & TYLER, L. (1980). The temporal structure of spoken language comprehension. *Cognition*, **8**, 1-72.
- MCCLELLAND, J. L., ST. JOHN, M., & TARABAN, R. (1989). Sentence comprehension: A parallel distributed processing approach. *Language & Cognitive Processes*, **4**, 287-336.
- PRITCHETT, B. L. (1992). *Grammatical competence and parsing performance*. Chicago: University of Chicago Press.
- RAYNER, K., SERENO, S., MORRIS, R., SCHMAUDER, R., & CLIFTON, C., JR. (1989). Eye movements and on-line language comprehension processes. *Language & Cognitive Processes*, **4**, 21-50.
- SCHMAUDER, A. R., & EGAN, M. C. (1995, November). *Verb subcategorization information, argument fit, and on-line sentence processing*. Poster presented at the 36th annual meeting of the Psychonomic Society, Los Angeles.
- SCHMAUDER, A. R., & EGAN, M. C. (1996, March). *On the impact of semantic fit and verb bias on the resolution of temporary syntactic ambiguity*. Poster presented at the ninth annual CUNY Conference on Human Sentence Processing, New York.
- SCHÜTZE, C. T. (1995). PP attachment and argumenthood. In C. T. Schütze, J. B. Granger, & K. Broihier (Eds.), *Papers on language processing and acquisition* (MIT Working Papers in Linguistics, No. 26, pp. 95-151). Cambridge, MA: MIT, Department of Linguistics.
- SCHÜTZE, C. T., & GIBSON, E. (1996, January). *Argument structure in English PP attachment*. Paper presented at the meeting of the Linguistic Society of America, San Diego.
- SPIVEY-KNOWLTON, M., & SEDIVY, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, **55**, 227-267.
- TARABAN, R., & MCCLELLAND, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory & Language*, **27**, 597-632.
- TRAXLER, M., & PICKERING, M. (1995, March). *Evidence against statistical parsing*. Poster presented at the eighth annual CUNY Conference on Human Sentence Processing, Tucson, AZ.
- TRUESWELL, J. C., & TANENHAUS, M. K. (1994). Toward a lexicalist framework of constraint-based syntactic ambiguity resolution. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 155-180). Hillsdale, NJ: Erlbaum.
- TRUESWELL, J. C., TANENHAUS, M. K., & KELLO, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 528-553.

NOTE

1. Schütze and Gibson (1996) reported a failure to replicate the Clifton et al. (1991) finding of an overall VP attachment advantage, using a word by word self-paced reading technique. They modified the Clifton et al. materials so that they were matched item by item in length and frequency rather than being matched over items in mean length and frequency, and so that all the items (not just 12 of 16) passed their explicit tests for argumenthood. We suspect that the main reason for the failure to replicate was the use of the word by word technique. Clifton et al. briefly reported results of two self-paced reading experiments using their materials, showing that the VP attachment advantage was highly dependent on how the presentations were segmented. Presenting regions smaller than the entire PP reduced or eliminated the effect. The fact that the effect was robustly obtained with eyetracking measures

suggests that reading strategies induced by word by word presentation eliminated an effect that is present in normal reading.

APPENDIX A

Experimental Materials in Experiments 1-2

Division into analysis regions for Experiment 2 are indicated by ^. Line breaks as they appeared in Experiment 2 are indicated by *. The letter X indicates the position of the phrase that generated the four versions of each sentence, which are shown in sequence after the sentence, followed by their plausibility ratings in parentheses: (a) Argument, high plausibility; (b) Argument, low plausibility; (c) Adjunct, high plausibility; (d) Adjunct, low plausibility.

1. Margaret^ began the letter^X,^ but she never^ finished^ * writing it.^

with a salutation (2.47); with a profanity (3.00); with a deep sigh (2.29); with a big shriek (3.35)

2. The wealthy investor^ paid ten thousand dollars^ *X^ and never missed^ the money.^

for a ski vacation (2.35); for free samples (4.71); for reasons of conscience (2.41); for the heck of it (4.29)

3. The coach^ discussed a new technique^X^ but finally^ * decided to stick with the old one.^

with the players (1.12); with the tailor (3.76); with enthusiasm (1.35); with sympathy (3.41)

4. Dr. Williams will vaccinate^ a Christian Scientist^X^ just * to prove^ the value^ of scientific medicine.

against measles (1.83); against backache (3.75); against her will (1.88); against all odds (3.29)

5. Little Willie^ had to ask his mother^X^ before he^ * could give us an answer.^

about the new puppy (1.76); about the empty cup (2.94); about a dozen times (1.94); about twice a second (3.35)

6. The people who lived near Love Canal^ blamed the toxic * waste dump^X^ but they never had^ enough resources^ * to sue.^

for their leukemia (1.41); for their hairdos (4.35); for several years (1.94); for a few moments (3.82)

7. His bad hearing^ exempted the young man^X^ but he tried^ * to enlist^ anyway.^

from the draft (1.76); from the shots (3.76); from the start (2.47); from the onset (3.35)

8. Maryellen^ begged her boyfriend^X,^ but it didn't do^ * any good.^

for a commitment (1.53); for a safety pin (2.76); for a whole week (1.94); for a millisecond (3.65)

9. The bridal consultant^ arranged the ceremony^X^ so the * bride felt^ confident^ about how it would all turn out.^

with a minister (1.65); with a plumber (4.12); with a smile (2.00); with a shrug (3.18)

10. The soldiers^ loaded the ammunition cartons^X^ * during the war^ in the desert.^

on an eighteen-wheeled truck (1.29); on a sterile operating table (4.12); on their routine assignment (2.06); on their special vacation (3.82)

11. The efficient repairwoman^ took the worthless TV set^ *X^ but the antenna is^ still here.^

to the repair shop (1.71); to the opera (4.76); to my great relief (2.47); to my utter chagrin (3.82)

12. The doctor^ lectured the obese patient^ X^ even * though he didn't think^ it would do any good.^

about heart disease (1.24); about canine obedience (3.76); about half an hour (2.12); about twelve hours (3.12)

13. I will contribute money in the spring. to the charitable foundation (1.06); to Ian's overfunded campaign (3.88); to the extent of my resources (1.94); to the limit of my endurance (3.06)

14. Your mother warned you and your father did too. about strange men (1.35); about bunny rabbits (4.12); about forty times (1.59); about two times (2.82)

15. The students questioned the English professor during office hours. about the papers (1.29); about physiology (4.24); about nine times (2.24); about one second (4.00)

16. We debated each other and still never convinced one another. about the decision (1.65); about the hamburger (2.82); about twenty times (2.29); about zero times (4.12)

17. Even miserly old Mrs. Wilson will donate clothing and furniture when times are as bad as now. to the Salvation Army (1.06); to the potluck dinner (3.88); to my amazement (2.18); to my complete awe (3.35)

18. The political organization tried to inform the voters during election season. about the ballot issues (1.53); about a fine restaurant (4.18); about twice every month (2.65); about a minute too late (3.18)

19. The local florist delivers flowers every Saturday afternoon. to my church (1.65); to my factory (1.94); to my delight (2.18); to my disdain (3.47)

20. The teacher really excited the students but couldn't ever get the principal's support. about a field trip (1.53); about more homework (2.53); about twice a year (2.76); about once a second (3.41)

APPENDIX B

Appendix B presents the adjusted mean reading times for Experiments 1 and 2, using rated plausibility of individual sentences as the covariate in the analysis of covariance from which the adjusted times are extracted. Note that the variable of plausible versus implausible is simply a discrete version of the variable used as the covariate, so that its effect largely disappears after the covariance adjustment.

Table B1

Argument Status	Plausibility		<i>M</i>
	Plausible	Implausible	
Argument	53.8	50.0	51.9
Adjunct	54.1	58.3	56.2
<i>M</i>	53.9	54.2	54.1

Note—*SE* of each cell mean is 1.80.

Table B2
Mean Adjusted Reading Times, Average of FF+1 and FF+2, Experiment 2

Argument Status	Plausibility		<i>M</i>
	Plausible	Implausible	
Argument	227	231	229
Adjunct	240	256	248
<i>M</i>	233	243	238

Note—*SE* of each cell mean is 7.17.

(Manuscript received June 11, 1996;
revision accepted for publication July 15, 1997.)