Journal of Memory and Language **45,** 225–258 (2001)
doi:10.1006/jmla.2001.2773, available online at http://www.academicpress.com on IDEAL®

# Reanalysis in Sentence Processing: Evidence against Current Constraint-Based and Two-Stage Models

Roger P. G. van Gompel and Martin J. Pickering

*University of Glasgow, Glasgow, Scotland, United Kingdom*

and

Matthew J. Traxler

*University of South Carolina*

Two eye-tracking experiments investigated processing of VP–NP attachment ambiguities. Experiment 1 tested sentences in which there was an initial bias toward VP attachment. Readers experienced more difficulty when semantic information disambiguated the sentences to NP attachment than when it disambiguated them to VP attachment or when it was consistent with either analysis. Experiment 2 tested sentences in which there was no initial bias toward either VP or NP attachment. Readers experienced more difficulty when semantic information disambiguated the sentences to NP attachment or VP attachment than when it was consistent with either analysis. We argue that these results challenge theories that assume a competition mechanism, such as constraint-based lexicalist accounts (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; McRae, Spivey-Knowlton, & Seidenberg, 1998; Spivey-Knowlton & Sedivy, 1995) and fixed-choice two-stage models (e.g., Frazier, 1987). We interpret the results in terms of the unrestricted race model (cf. Traxler, Pickering, & Clifton, 1998). 

*Key words:* sentence processing; syntactic ambiguity resolution; reanalysis; competition.

Two competing and largely incompatible classes of account dominate current sentence processing research. According to one class, the processor computes syntactic analyses serially, in two stages. In the first stage, it draws on a restricted range of information sources to construct an initial analysis. During the second stage, it accesses other sources of information, which may sometimes cause it to abandon its initial analysis and compute another. The best known account of this class is the *garden-path* model, in which the processor makes initial decisions on the basis of strategies defined in terms of syntactic information alone and uses thematic information in the second stage (e.g., Ferreira & Clifton, 1986; Ferreira & Henderson, 1990; Frazier, 1979, 1987; Frazier & Rayner, 1982; Rayner, Carlson, & Frazier, 1983). Other two-stage accounts assume that the processor uses thematic information immediately (Abney, 1989; Crocker, 1995; Pritchett, 1992; see Clifton, Speer, & Abney, 1991).

The second class of model assumes that the processor can activate multiple analyses in parallel. It employs both syntactic and nonsyntactic information in a single stage to foreground one analysis, but other analyses remain activated (e.g., MacDonald, 1994; MacDonald, Pearlmutter, & Seidenberg, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Spivey-Knowlton & Sedivy, 1995; Trueswell, 1996; Trueswell, Tanenhaus, & Garnsey, 1994; Trueswell, Tanenhaus, & Kello, 1993; Taraban & McClelland, 1988; Tyler & Marslen-Wilson,

We thank Steven Frisson, Simon Garrod, Simon Liversedge, Don Mitchell, Robin Morris, Keith Rayner, Christoph Scheepers, and an anonymous reviewer for useful suggestions and discussions and Keith Edwards for technical support. Special thanks to Chuck Clifton for the use of his software. Roger van Gompel acknowledges an ESRC Postgraduate Studentship and a University of Glasgow Postgraduate Studentship. Martin Pickering acknowledges support of ESRC Research Grant R000234542. Matt Traxler acknowledges support from the CLASS and SPAR programs at the University of South Carolina.

Address correspondence and reprint requests to Roger van Gompel, Department of Psychology, University of Dundee, Dundee DD1 4HN, Scotland, UK. Fax: +44 1382 229993. E-mail: r.p.g.vangompel@dundee.ac.uk.



0749-596X/01 $35.00

1977). Current models of this type are known as *constraint-based lexicalist* models.

Current data cannot satisfactorily distinguish these accounts. Most research has attempted to determine whether all available information is used during the very earliest stages of processing. Some studies suggest that the processor initially ignores subcategorization preferences, semantic plausibility, and felicity with respect to discourse context (e.g., Ferreira & Henderson, 1990; Ferreira & Clifton, 1986; Mitchell, 1987; Mitchell, Corley, & Garnham, 1992; Rayner et al., 1983; Rayner, Garrod, & Perfetti, 1992). However, other studies find evidence that it uses one or more of these sources of information immediately (Altmann, Garnham, & Dennis, 1992; Altmann, Garnham, & Henstra, 1994; Altmann & Steedman, 1988; Britt, 1994; Britt, Perfetti, Garrod, & Rayner, 1992; Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Trueswell, 1996; Trueswell et al., 1993, 1994). In order to account for the latter findings, proponents of two-stage accounts claim that the second stage may begin extremely quickly and that this latter set of studies tapped processes occurring during the second stage. They suggest that currently available techniques may be insensitive to processes occurring in the first stage (e.g., Clifton & Ferreira, 1989). On the other hand, proponents of one-stage accounts argue that studies that fail to find early effects used weak manipulations.

### *The Source of Processing Difficulty: Reanalysis vs Competition*

Let us therefore consider whether the accounts can be distinguished in other ways. Apart from strategies and information sources used during initial processing, the major difference between the models concerns the mechanisms underlying ambiguity resolution.

*Reanalysis models*. Two-stage models are, in general, serial models. The processor initially selects one analysis, say analysis A. If this analysis subsequently becomes impossible (or unlikely), the processor reanalyzes, thereby adopting analysis B, and disruption ensues (e.g., Frazier & Rayner, 1982; Rayner et al., 1983). If the initial analysis is consistent with later information, there is no need for reanalysis. It does not matter whether other analyses are possible, because they have not been considered.

An important distinction within two-stage models is between theories that assume that the processor (of any speaker of the language) always adopts the same analysis of a particular syntactic ambiguity and theories that do not. We refer to the former as *fixed-choice two-stage models* and the latter as *variable-choice two-stage models*. The garden-path theory (Frazier, 1979, 1987) is the best known example of the former. In this model, the processor always follows the structural principles of minimal attachment and late closure in determining which analysis to adopt. The nonpreferred analysis is never adopted.

Several other models can also be considered fixed-choice models. These models have in common that they assume that the processor always uses the same parsing principles for any particular structure and that there is no random noise that may cause a nonpreferred analysis to be selected. For example, in thematic accounts such as Abney's (1989), readers should always initially adopt argument attachment analyses. Another example of a fixed-choice model is referential theory (Altmann & Steedman, 1988; Crain & Steedman, 1985; Ni, Crain, & Shankweiler, 1996). Although the processor initially proposes multiple analyses in parallel, it rapidly selects the analysis that requires fewest unsatisfied presuppositions. Analyses that contain more presuppositions are never selected. In the case of ambiguities involving a definite noun phrase, this leads to the selection of an analysis involving a simple noun phrase over a complex noun phrase in a null context or a context mentioning one entity of the type referred to by the noun phrase, and a selection of the analysis involving the complex noun phrase in a context mentioning more than one such entity. (Note that referential theory and thematic accounts, unlike garden-path theory, do not necessarily make predictions for all syntactic ambiguities.)

There are few clearly articulated variable-choice models. The main exception is the tuning hypothesis, in which parsing decisions are determined by the frequency with which the alternative analyses are used in the language

(Mitchell, Cuetos, Corley, & Brysbaert, 1995; see also Brysbaert & Mitchell, 1996; Mitchell & Cuetos, 1991). Assuming that speakers may differ in their exposure to different analyses (Cuetos, Mitchell, & Corley, 1996), the tuning hypothesis predicts that the analysis that different people assign to a particular sentence need not be the same. However, given that frequency is the only type of information readers use during initial analysis, any given person will have a fixed preference for a particular sentence type.

Traxler, Pickering, and Clifton (1998) proposed another variable-choice model, which we shall call the *unrestricted race model* (see also Traxler, Pickering, Clifton, & Van Gompel, 2000; Van Gompel, Pickering, & Traxler, 2000; and Lewis, 1999; McRoy & Hirst, 1990, for similar proposals). As in constraint-based theories, there is no restriction on the sources of information that can provide support for the different analyses of an ambiguous structure; hence it is *unrestricted*. In the model, the alternative structures of a syntactic ambiguity are engaged in a race, with the structure that is constructed fastest being adopted. Although it is assumed that the processor attempts to construct multiple analyses in parallel, only a single analysis is adopted. If this analysis is inconsistent with later information, the processor has to reanalyze, and processing difficulty occurs. Thus, the unrestricted race model is a two-stage reanalysis model. It is also a variable-choice model, because the initial analysis of a syntactic ambiguity is affected both by individual differences (as in the tuning hypothesis) and by syntactic and nonsyntactic characteristics of the sentences. Because the strength of these sources of information may differ for each individual sentence, preferences may differ for each sentence read by a particular person. Indeed, we assume that any unrestricted two-stage account is a variable-choice model.

In support of this model, Traxler et al. (1998) conducted three eye-tracking experiments on adjunct attachment ambiguities involving relative clauses. They compared globally ambiguous sentences with sentences that were disambiguated toward one or other analysis. In one experiment, they investigated sentences such as in (1). In (1a) the relative clause can plausibly be attached to either NP1 (*the son*) or NP2 (*the driver*), whereas in (1b) and (1c) only one analysis is plausible:

(1a) The son of the driver that had the moustache was pretty cool. (ambiguous)

(1b) The car of the driver that had the moustache was pretty cool. (NP2 attachment)

(1c) The driver of the car that had the moustache was pretty cool. (NP1 attachment)

Traxler et al. observed that total reading times for the critical word (*moustache*) were longer in the disambiguated conditions (1b) and (1c) than in the ambiguous condition (1a). Another experiment showed a similar pattern for relative clause ambiguities that were disambiguated by the gender of a reflexive pronoun (*himself/herself*), but the effect occurred somewhat earlier, on regressions from the words immediately following the disambiguating word.

Because both disambiguated conditions produced difficulty relative to the ambiguous condition, these results suggest that readers did not always adopt the same analysis. Instead, they suggest that the processor adopted the NP1 analysis on some proportion of trials, resulting in difficulty when the sentence was disambiguated toward NP2 attachment; and the NP2 analysis on another proportion, resulting in difficulty when the sentence was disambiguated toward NP1. Hence, these results are difficult to reconcile with fixed-choice models.

However, the results only provide evidence for variable-choice accounts of adjunct ambiguities. It may not be possible to generalize from adjunct ambiguities to argument ambiguities, as many theories make clear distinctions between the processing of adjuncts and arguments (Abney, 1989; Frazier & Clifton, 1996; Liversedge, Pickering, Branigan, & Van Gompel, 1998; Pritchett, 1992; Schütze & Gibson, 1999). In particular, construal theory (Frazier & Clifton, 1996) makes a distinction between primary and nonprimary phrases, which corresponds roughly to the distinction between arguments and adjuncts. Frazier and Clifton claimed that ambiguities involving nonprimary relations such as relative clause attachment ambiguities are resolved differently from ambiguities involving primary relations. Ambiguities that involve nonprimary phrases are resolved by the principle of construal: Nonprimary relations are associated with all possible attachment sites within the current thematic domain (defined in terms of the most recent theta assigner), and the final choice of attachment is made between these different sites in an unrestricted manner. Therefore, it is a variable-choice model with respect to ambiguities that involve nonprimary phrases. However, construal is a fixed-choice model with respect to ambiguities that involve at least one primary phrase. If a sentence is ambiguous between a primary and non-primary relation, the processor always adopts the primary relation. Furthermore, if a sentence is ambiguous between two structures that both involve primary relations, construal predicts that the processor always adopts the structurally simplest analysis (determined by minimal attachment and late closure). Traxler et al.'s (1998) data are consistent with a variable-choice model of nonprimary relations, but a fixed-choice model of primary relations. In order to see if the predictions of the unrestricted race model hold generally, we need to investigate ambiguities that involve primary relations.

*Competition models*. In most constraint-based lexicalist models, syntactic ambiguity resolution involves a mechanism of *competition*. Competition is one source of processing difficulty in such models (MacDonald et al., 1994; McRae et al., 1998; Spivey & Tanenhaus, 1998; Spivey-Knowlton & Sedivy, 1995; Tabor, Juliano, & Tanenhaus, 1997; Tabor & Tanenhaus, 1999; Tanenhaus & Trueswell, 1995). When alternative analyses of an ambiguity have near-equal activation, competition is particularly strong and severe processing difficulty is predicted. Spivey-Knowlton and Sedivy (1995) proposed "Rather than interpreting elevated reading times as evidence that an initial analysis has been identified as incorrect, and that a new structure is required, our current approach suggests that processing delays are a manifestation of direct competition between opposing alternatives [. . .] Near equal activation levels will result in lengthy competition, hence greatly slowed reading times at the point of ambiguity" (p. 260).

Constraint-based accounts assume that syntactic ambiguity resolution is closely related to lexical ambiguity resolution (MacDonald et al., 1994; see also Spivey-Knowlton & Sedivy, 1995; Trueswell et al., 1993, 1994; Garnsey et al., 1997). Research on lexical ambiguity resolution shows that in neutral contexts, reading time for balanced ambiguous words (i.e., words with two meanings of approximately equal frequency) is longer than reading time for unambiguous control words (Rayner & Duffy, 1986; Duffy, Morris, & Rayner, 1988). By contrast, Duffy et al. (see also Binder & Rayner, 1998; Rayner, Pacht, & Duffy, 1994) showed that when balanced ambiguous words are disambiguated by prior context, they are processed as quickly as unambiguous control words. Biased ambiguous words (i.e., words with two meanings of greatly unequal frequency) are processed slowly when preceding context disambiguates the word toward its less frequently occurring meaning. This is the so-called subordinate-bias effect (Rayner et al., 1994). Thus, lexical processing is slow when neither context nor frequency provides good evidence about which interpretation should be favored or when different sources of information support different meanings. This suggests that competition occurs in lexical ambiguity resolution when two meanings of an ambiguous word are about equally activated by the constraints (frequency and context). Because syntactic ambiguity resolution is assumed to be similar to lexical ambiguity resolution, constraint-based theories predict competition to occur in syntactic ambiguity resolution as well. For instance, MacDonald et al. (1994) claimed that "[. . .] X-bar structures will each be partially activated to some degree, depending on the strength of their support from the syntax and from other parts of the system (e.g., their associated argument structures), and they will compete with each other just as other representations at the same level do within a single lexical entry" (p. 687). In contrast, two-stage models draw a clear distinction between syntactic and lexical ambiguity resolution (Frazier, 1987, 1995; Frazier & Clifton, 1996; Rayner & Morris, 1991; Traxler et al., 1998). Proponents of such a distinction usually assume that word meanings are accessed in parallel, which can give rise to competition effects, whereas syntactic structures are constructed in a serial fashion, so that no competition can occur.

Competition models predict that in syntactic ambiguity resolution, there are two cases where competition arises. In the first case, constraints favor analysis A as soon as the ambiguity arises. But constraints that come in at one of the following words favor analysis B instead (e.g., because they provide disambiguating information). Thus, this case involves a change in preferences as a result of constraints that emerge at different words. This leads to near-equal activation of the two analyses and results in competition, which causes processing difficulty. Note that a two-stage model in which analysis A is initially adopted predicts reanalysis in such cases.

In the second case, constraints support analyses A and B approximately equally as soon as the ambiguity arises, and later words do not favor either analysis. This also causes competition. In this case, difficulty caused by competition has no equivalent within two-stage models. They predict no reanalysis and therefore no difficulty in such cases. Hence, investigation of this second case should allow us to distinguish between reanalysis- and competition-based models.

An important issue with respect to ambiguities that belong to the second case is the time course of competition. When multiple analyses remain equally supported during several words, for how long will the analyses remain activated in parallel before one analysis is selected and competition is resolved? Some researchers have argued that competition continues for a very long time. A clear example of such a view is found in MacDonald, Just, and Carpenter (1992), who discussed sentences such as (2a) and (2b):

(2a) The experienced soldiers warned about the dangers before the midnight raid.

(2b) The experienced soldiers spoke about the dangers before the midnight raid.

In both sentences, the main clause analysis is the ultimately correct analysis. However, (2a) is ambiguous between a main clause and reduced relative analysis until the period is reached, whereas (2b) is unambiguous because *spoke* cannot be a past participle. MacDonald et al. claimed that if both analyses are activated in parallel in (2a), this sentence should be harder to read than (2b), where only one analysis is possible. Importantly, they claimed that it should be possible to find such effects in the region *before the midnight*, well downstream from the point where the ambiguity arises (at *warned*). Hence, they predicted very long-lasting competition.

Some recent computational models also favor long-lasting competition, but they claim that competition gradually decreases. The *competition-integration* model of McRae et al. (1998) and Spivey and Tanenhaus (1998) is an example of such a model (see Tabor & Tanenhaus, 1999, for a model that makes similar predictions). In this computational model, the different analyses of an ambiguity receive support from various sentence processing constraints. At each word, the analysis that is supported most strongly gradually becomes more activated due to cycles of forward and backward activation between the analyses and the constraints. These cycles continue until one analysis reaches a certain threshold level (determined by a free parameter). The longer it takes to reach this threshold, the stronger the competition. Next, the processor moves to the following word. At this word, the final activation levels of the analyses at the previous word are treated as one of the constraints for the current word (the competition-integration model stipulates that this constraint has the same weight as all the new constraints combined). If the new constraints support two analyses about equally, the activation levels of the two analyses will initially become more equal again. But subsequently, cycles of forward and backward activation between the analyses and constraints will gradually increase the difference in activation levels until the threshold level is reached again and the processor moves to the next word. In the competition-integration model, it is possible to set the value of the threshold level such that the difference between the final activation levels at a word is greater than at the preceding word. As a result, competition will gradually decrease as the processor moves from one word to the next. However, competition will never completely disappear: Even the least activated analysis remains (somewhat) activated.

In fact, most constraint-based competition models are unclear about the precise time course of competition. However, to our knowledge, none of the these theories claim that the nonpreferred analysis is abandoned completely. In all theories, multiple analyses remain activated in parallel (though perhaps only weakly) at least as long as the sentence is ambiguous (e.g., McRae et al., 1998; Spivey & Tanenhaus, 1998; Spivey-Knowlton & Sedivy, 1995; Tabor et al., 1997; Tabor & Tanenhaus, 1999). Hence, they assume that competition is long-lasting.

### *Evidence for Competition*

In the past, a number of studies have claimed to find evidence for competition in syntactic ambiguity resolution. MacDonald et al. (1992) investigated sentences such as (2a) and (2b) and observed that reading times for *before the midnight* were longer in (2a) than in (2b). MacDonald et al. took this as evidence that readers retained both analyses in memory in the temporarily ambiguous condition, whereas only one analysis was available in the unambiguous condition. This can be interpreted as competition between the reduced relative and the main clause analysis in the ambiguous condition. Two further experiments replicated these results. These findings clearly pose difficulty for reanalysis models, which claim that only one analysis is pursued. However, this study was subsequently criticized by Pearlmutter and MacDonald (1995). They conducted a plausibility study on MacDonald et al.'s materials and showed that the ambiguous conditions were rated as less plausible than the unambiguous conditions. They argued that this probably explained the relatively long reading times in the ambiguous condition. Therefore, the MacDonald et al. study does not provide convincing evidence for competition in syntactic ambiguity resolution.

In another study, MacDonald (1994) attributed longer reading times in the region *captured in the coup* in (3a) than in *fought in the coup* in (3b) to competition between constraints.

(3a) The ruthless dictator captured in the coup was hated throughout the country.

(3b) The ruthless dictator fought in the coup was hated throughout the country.

She suggested that the *main clause* analysis (as in, e.g., *The ruthless dictator captured the city*) in (3a) and (3b) is activated because it occurs most frequently or is least complex. In (3a), *captured in the coup* is barely consistent with the main clause analysis, so it is deactivated, and the (ultimately correct) *reduced relative* analysis gets more activation. This leads to competition. But in (3b), *fought in the coup* is compatible with the main clause analysis. Thus there is very little competition here. In this region, (3a) was harder than (3b) in relation to unambiguous controls.

In another study, McRae et al. (1998) observed processing difficulty when plausibility information was inconsistent with the main clause analysis and claimed that this was due to competition between the main clause and reduced relative analysis. However, the effects in MacDonald (1994) and McRae et al. (1998) could equally well have been due to reanalysis: The processor may initially have adopted the main clause analysis, but then started to reanalyze when evidence rendered this analysis unlikely. Hence, these data do not distinguish between reanalysis and competition.

### *The Processing of VP–NP Attachment Ambiguities*

Because previous studies have not satisfactorily discriminated between reanalysis and competition, we conducted a new study to investigate this issue. In this study, we investigated *VP–NP attachment ambiguities* like (4) and (5):

(4a) The hunter killed the dangerous poacher with the rifle not long after sunset. (ambiguous)

(4b) The hunter killed the dangerous leopard with the rifle not long after sunset. (VP attachment)

(4c) The hunter killed the dangerous leopard with the scars not long after sunset. (NP attachment)

(5a) The hunter killed only the poacher with the rifle not long after sunset. (ambiguous)

(5b) The hunter killed only the leopard with the rifle not long after sunset. (VP attachment)

(5c) The hunter killed only the leopard with the scars not long after sunset. (NP attachment)

The prepositional phrase *with the rifle/scars* attaches either to *killed* (the *VP attachment* analysis), whereby the hunter killed with the rifle or scars, or to *leopard/poacher* (the *NP attachment* analysis), whereby the leopard or poacher had the rifle or scars. Experiment 1 employed sentences like (4); Experiment 2 employed sentences like (5).

These sentences are produced by making two manipulations. The first concerns the plausibility of the two analyses. In the ambiguous conditions [(4a) and (5a)], both analyses are plausible; in the VP conditions [(4b) and (5b)], only the VP attachment is plausible; and in the NP conditions [(4c) and (5c)], only the NP analysis is plausible. We assessed the plausibility manipulation in plausibility pretests. The second manipulation is the initial preference for VP versus NP attachment (independent of the plausibility manipulation). We assessed this preference in two ways: attachment preferences when plausibility was carefully controlled [as in (4a) and (5a)], and completion preferences for fragments ending just before *the rifle* or *the scars*. These pretests demonstrated that the initial preference in (4) was for VP attachment, whereas (5) had no clear initial preference for VP or NP attachment. We now outline the predictions of different theories for the processing of such sentences.

*Constraint-based theories*. Constraint-based theories predict strong competition in (4c). At *with the*, the VP attachment preference favors the VP analysis. However, at *scars*, the plausibility disambiguation favors the NP analysis. This reversal of preferences leads to strong competition between the VP and NP analysis. In the VP attachment condition (4b), no competition is predicted, because the VP attachment preference at *with the* is consistent with the plausibility disambiguation. In the ambiguous condition (4a), the VP attachment preference biases toward the VP analysis at *with the*, but plausibility does not favor either analysis. Hence, the degree of competition should be intermediate between (4b) and (4c). However, if the activation of the VP analysis is very high when the processor reaches *rifle*, this may outweigh any competition effects due to equal plausibility of the two analyses, and therefore it might, in practice, be impossible to distinguish (4a) from (4b) [but (4a) should not be easier than (4b)].

Competition-based theories also predict competition in (5a), the ambiguous condition for the balanced ambiguity. Because the materials are balanced, neither the VP nor NP analysis is preferred when the ambiguity arises (at *with*). This results in competition at this point. Assuming that competition is relatively long-lasting, the analyses will continue competing at the following words (*the rifle*) because these words do not provide any biasing information. Thus, (5a) should be hard to read at the words following *with*. However, if competition decreases very rapidly, processing difficulty may become negligible after a few words. We return to this issue in Experiment 2.

Competition should be much weaker in the VP and NP disambiguated conditions (5b) and (5c). Competition in these sentences occurs at *with the* because the ambiguity is balanced. But of course the same amount of competition occurs in (5a), as the sentences do not differ at this point. At disambiguation, plausibility information provides a strong constraint in favor of one of the analyses (while other constraints do not favor either analysis). Thus, processing of the disambiguated conditions (5b) and (5c) should be easier than the ambiguous condition (5a). No difference is predicted between the disambiguated conditions, because our pretests indicated that the plausibility bias was equally strong for the two conditions and there was no initial preference for one of the analyses.

*Fixed-choice two-stage theories*. Most two-stage theories predict that the processor always initially adopts the VP analysis in sentences like (4). According to garden-path theory, this follows from the principle of minimal attachment because it is assumed that the VP analysis employs fewer nodes in the phrase structure tree than the NP analysis (Frazier, 1979; Rayner et al., 1983).

Construal theory (Frazier & Clifton, 1996) claims that both VP and NP attachment are initially interpreted as primary relations (even though they may not be primary relations in their final interpretation), and therefore, the construal principle does not apply in such sentences. Hence, the processor bases its initial analysis on structural principles only. As in the garden-path model, the processor should always initially adopt the VP analysis as a result of the principle of minimal attachment.

Finally, referential theory (Altmann & Steedman, 1988; Crain & Steedman, 1985; Ni et al., 1996) claims that the processor adopts the analysis with the fewest unsatisfied presuppositions. In (4), the VP analysis involves a noun phrase (e.g., *the dangerous poacher*) which requires the processor to postulate the existence of a single dangerous poacher; whereas the NP analysis involves a complex noun phrase (e.g., *the dangerous poacher with the scars*), which requires the processor to postulate the existence of a set of dangerous poachers. As postulating a single dangerous poacher requires fewer presuppositions than postulating a set of dangerous poachers, the VP analysis is adopted.

Most two-stage accounts make the same predictions for (5) as (4). The exception is the referential theory as interpreted by Ni et al. (1996). Ni et al. argued that in referential theory, the inclusion of the word *only* makes the NP attachment preferred. If *only* precedes a NP such as *poacher* in (5), the processor assumes more than one poacher. As a result, it analyzes the following prepositional phrase as a modifier which specifies the right poacher and therefore adopts the NP analysis. They found some evidence from eye-tracking and self-paced reading in favor of this account. However, other researchers have not found evidence for such strong effects of *only* on syntactic ambiguity resolution (Clifton, Bock, & Radó, 2000; Paterson, Liversedge, & Underwood, 1999).

*Variable-choice two-stage theories*. Variable-choice models differ from fixed-choice models with respect to the proportion of times that the VP or NP analysis is initially chosen. The tuning hypothesis assumes that the initial attachment preference for sentences such as (4) and (5) is determined by how often readers have encountered that particular structure. Mitchell et al. (1995) noted that such frequency information can be represented at different levels of linguistic representation. They made a distinction between fine-grained frequency information, when information about the frequency of syntactic structures is stored for each individual lexical item, and coarse-grained frequency information, when it is stored for larger syntactic structures (e.g., for all VP–NP attachment structures). In a number of articles (e.g., Brysbaert & Mitchell, 1996; Cuetos, Mitchell, & Corley, 1996; Mitchell et al., 1995), Mitchell and colleagues have tentatively favored the coarse-grained variant of the tuning hypothesis. This variant predicts that a reader's preference for VP or NP attachment is determined by the number of times a reader has previously encountered VP and NP attachment in V–NP–PP structures. The particular verb, noun phrase, or preposition that is used is not taken into account. In particular, as *only* constitutes part of the noun phrase, its presence should not affect preferences. Therefore, the same preference should occur in (4) and (5). Some corpus studies have found that NP attachment is more frequent than VP attachment in V–NP–PP structures (Collins & Brooks, 1995; Hindle & Rooth, 1993; Merlo, Crocker, & Berthouzoz, 1997). If these results are representative of the V–NP–PP structures that our participants encountered, the tuning hypothesis predicts that most readers initially adopt the NP analysis. However, it is possible that some participants may have encountered more VP than NP attachments. Consequently, reanalysis may occur on some trials in both the VP and NP attachment condition, and both may be more difficult than the ambiguous condition.

In the unrestricted race model, preferences may differ for each sentence that is read by a particular person because preferences are affected both by item characteristics and by individual factors. Thus, readers may adopt either the VP or NP analysis initially. In (4), there is a strong initial bias for the VP analysis. The processor takes this bias into account and therefore nearly always adopts the VP analysis initially. Thus, this theory predicts that readers will experience difficulty when this analysis becomes implausible, as in (4c), but that little or no difficulty should occur when the NP analysis becomes implausible, as in (4b). No processing difficulty should occur in the ambiguous sentence (4a), because the initial analysis is always plausible.

In contrast, because there is no bias for either VP or NP attachment in (5), the model predicts that the processor will adopt each analysis on about half of the trials initially. Hence it makes a novel prediction for (5): In the VP condition (5b) and the NP condition (5c), the processor will be forced to reanalyze on those trials where it adopted the implausible analysis; but in the ambiguous condition (5a), it will never have to reanalyze, because either analysis is plausible. The ambiguous condition (5a) should be easier than (5b) and (5c) (which should not differ from each other). Hence, the unrestricted race model makes essentially the opposite predictions to constraint-based theories for (5). It also makes different predictions from fixed-choice models for sentences like (5).

In all two-stage theories, the processor only reanalyzes if the initial analysis becomes implausible or unlikely. It should not notice whether the nonselected analysis is implausible so long as the selected analysis remains plausible. Thus, if it initially adopts the VP analysis, it should experience difficulty with (4c). But (4a) and (4b) should not differ in difficulty. If it initially adopts the NP analysis, it should experience difficulty with (4b), but not with (4a) or (4c). The same argument applies for the sentences in (5). The crucial prediction is that the globally ambiguous sentences [(4a) and (5a)] should not be more difficult to process than the disambiguated sentences.

*Summary*. Constraint-based theories predict competition either when two constraints support different analyses [as in (4c)] or when both constraints support two analyses to an equal degree [as in (5a)]. Fixed-choice models (e.g., garden-path theory, construal theory, and referential theory) predict difficulty when the disambiguation is inconsistent with an initial choice of analysis: Either the NP condition or the VP condition should cause difficulty. The ambiguous condition should be as easy as the disambiguated condition that produces least difficulty, but no easier. A variable-choice account such as the unrestricted race model predicts that the NP condition (4c) should be harder than the ambiguous condition (4a) and the VP condition (4b); and the ambiguous condition (5a) should be easier to process than either of the disambiguated conditions [(5b)–(5c)].

## EXPERIMENT 1

We monitored participants' eye movements as they read VP–NP attachment ambiguities like [(4a)–(4c)]. These ambiguities had an initial bias toward VP attachment. Note that the predictions of most theories do not differ for these sentences. Although the experiment rigorously controlled for plausibility and introduced a novel ambiguous condition, its main purpose was to determine patterns of eye movements for these sentences so that precise predictions could be made for the more critical sentences [(5a)–(5c)] that were employed in Experiment 2.

### *Participants*

Thirty-six participants were paid to take part in the eye-tracking experiment. All of the participants in all parts of the study were native-English-speaking students at the University of Glasgow and had normal or corrected-to-normal vision. Some had taken part in other eye-tracking experiments. No one participated in more than one experiment or pretest reported in this article. Two further participants were excluded, because they answered more than 25% of the statements incorrectly.

*Items*

We constructed 30 items like (4), repeated below (see Appendix):

(4a) The hunter killed the dangerous poacher with the rifle not long after sunset. (ambiguous)

(4b) The hunter killed the dangerous leopard with the rifle not long after sunset. (VP attachment)

(4c) The hunter killed the dangerous leopard with the scars not long after sunset. (NP attachment)

The VP attachment condition was semantically disambiguated to the VP analysis; the NP attachment condition was semantically disambiguated to the NP analysis; the Ambiguous condition was semantically ambiguous between the two analyses (see plausibility pretest). All items consisted of a subject noun phrase, a verb, an object noun phrase (containing an adjective as well as a noun), a prepositional phrase (consisting of the preposition *with* and a noun phrase), and a sentence-final adverbial phrase. The direct object and the noun phrase in the prepositional phrase contained the definite determiner *the*. On the VP analysis, the prepositional phrase was interpreted as an instrument of the verb; on the NP analysis, it was interpreted as an attribute of the object noun. The object noun phrase and the noun within the prepositional phrase differed between conditions. The object in the NP attachment condition was the same as the object in the VP attachment condition for half of the items, and the same as the object in the ambiguous condition for the other half. The noun within the prepositional phrase was the same in the ambiguous and VP attachment conditions, whereas the NP attachment condition had a different noun. The total length of the adjective plus noun was identical for each version of a given item, as was the length of the noun within the prepositional phrase. According to the Celex English database (Baayen, Piepenbrock, & Van Rijn, 1993), word-form frequency for the noun within the prepositional phrase did not differ between conditions ($F < 1$), nor did the logarithm of the word-form frequency ($F < 1$) (see Table 1).

*Pretests*

The items were subjected to three pretests. These tests confirmed that the Ambiguous condition was (semantically) plausible on both analyses, that the other conditions were plausible on one analysis but implausible on the other, and that the materials were biased (independent of plausibility) toward VP attachment.

*Plausibility pretest*. We carefully controlled our sentences for plausibility. Previous studies have often manipulated plausibility by intuition alone, with the result that VP attachment conditions have often included ambiguous sentences where the NP analysis is also plausible (e.g., *The spy saw the cop with binoculars but the cop didn't see him*, from Rayner et al., 1983). Additionally, even unambiguous sentences can differ in plausibility from each other, and such differences can impact reading times (e.g., Traxler & Pickering, 1996). We therefore ruled out the possibility that any effect was due to an irrelevant difference in plausibility. We also constructed the sentences so that plausibility was manipulated on a single word [*rifle* or *scars* in (4)].

The plausibility pretest investigated the plausibility of the different analyses. It began with 50 sets of items and reduced them to the final set of 30. We derived six sentences like (6a)–(6f) from each item set.

(6a) The hunter killed the dangerous poacher by using the rifle.

(6b) The hunter killed the dangerous leopard by using the rifle.

(6c) The hunter killed the dangerous leopard by using the scars.

(6d) The dangerous poacher with the rifle was killed by the hunter.

TABLE 1

Experiment 1: Item Characteristics

| | Condition | | |
|---|---|---|---|
| | Ambiguous | VP attachment | NP attachment |
| Plausibility pretest | | | |
| Plausibility VP-PP paraphrase | 6.1 | 6.1 | 1.5 |
| Plausibility NP-PP paraphrase | 5.5 | 2.1 | 5.9 |
| Frequency of noun in PP | 406 | 406 | 479 |
| Log frequency of noun in PP | 1.88 | 1.88 | 2.00 |
| Off-line preference task | | | |
| Percentage of VP attachment | 78.3 | 95.6 | 2.2 |
| Completion task | | | |
| Percentage of VP completions | 93.9 | 94.9 | 90.1 |

*Note*. Frequency is the number of occurrences of the word form in the Celex database (17.9 million words).

(6e) The dangerous leopard with the rifle was killed by the hunter.

(6f) The dangerous leopard with the scars was killed by the hunter.

These sentences were unambiguous paraphrases of item sets like (4a)–(4c), on the VP or the NP analysis. For the VP analysis (6a)–(6c), we replaced *with* with *by using*, *by means of* (two item sets) and *with the help of* (two item sets). We refer to these conditions as the *VP–PP paraphrase conditions*. For the NP analysis (6d)–(6f), we converted the original sentence into a passive. We refer to these conditions as the *NP–PP paraphrase conditions*. Notice that the VP–PP paraphrase of the NP attachment condition (6c) and the NP–PP paraphrase of the VP attachment condition (6e) derived from analyses that were constructed to have implausible interpretations. All other conditions were derived from analyses that were constructed to be plausible. Twenty-two participants rated how realistic they thought that the situations described by (6a)–(6f) were, on a 7-point scale, with 7 indicating *a very realistic situation* and 1 indicating *a very unrealistic situation*. Each participant rated all 300 sentences, so that all conditions of an item were rated by the same participants. This enabled us to make a reliable comparisons between the conditions of each individual item. At least 20 sentences appeared between two conditions of a single item. Half the participants received the items in one random order, and half received the exact reverse order of items.

We then selected the final 30 item sets. Table 1 shows the mean plausibility rating for each condition for each analysis. We conducted 2 (Paraphrase type: VP–PP vs NP–PP paraphrase) $\times$ 3 (Condition: ambiguous vs VP attachment vs NP attachment) ANOVAs, treating participants and items as random factors. These revealed a significant interaction between the factors: $F1(2, 42) = 436$, $p < .001$, $MS_e = .45$; $F2(2, 58) = 1239$, $p < .001$, $MS_e = .22$. For the purposes of the experiment, it was crucial that conditions (6c) and (6e) were implausible, but the other conditions were not; Table 1 shows that this was indeed the case. Planned comparisons showed that the two paraphrase types differed for both the VP attachment condition and the NP attachment condition (All $p$s $< .001$). Although the difference between the VP–PP and NP–PP paraphrases in the ambiguous condition was significant ($F1 = 9.71$, $p = .003$, $F2 = 27.6$, $p < .001$), it was numerically very small.

We also analyzed the absolute differences between the scores for the two paraphrase types in the three conditions. Planned comparisons revealed that this difference was much smaller in the ambiguous condition than in the VP condition and the NP condition (all $p$s $< .001$). Hence the plausibility difference was much less in the ambiguous conditions than in the disambiguated conditions.

*Off-line preference task.* This pretest investigated the bias of the items as they were used in the eye-tracking experiment. Eighteen participants received exactly the same sentences as in the eye-tracking experiment, with six participants seeing each list. Each item was followed by the two paraphrases of the item used in the plausibility pretest, balanced for order. Participants indicated whether the VP–PP or the NP–PP paraphrase best described the situation described by the sentence.

The crucial condition to test the bias of the items is the ambiguous condition. Given that the plausibility pretest found that both attachments were about equally plausible in this condition, a preference for either the VP or NP analysis must be the result of a bias in the items before the critical noun [i.e., *rifle* in (4a)], independent of plausibility. Testing a semantically ambiguous condition is a common method to test for syntactic preferences (e.g., Cuetos & Mitchell; 1988; De Vincenzi & Job, 1993; Frazier & Clifton, 1996; Gilboy, Sopena, Clifton, & Frazier, 1995; Hemforth, Konieczny, & Scheepers, 2000).

We predicted that the VP and NP conditions would normally be given VP and NP interpretations respectively. In other words, participants should avoid implausible interpretations. This would indicate that our plausibility manipulations were strong enough to eventually override any syntactic bias. Table 1 indicates that this was indeed the case (the two paraphrases always added up to 100%). Analyses of variance, treating Condition (ambiguous vs VP vs NP attachment) as a within-participants and within-items factor, revealed a significant effect of Condition: $F1(2, 34) = 436$, $p < .001$, $MS_e = 102$; $F2(2, 58) = 314$, $p < .001$, $MS_e = 236$. Planned comparisons revealed that the percentage of VP–PP attachment responses in the ambiguous condition was different from both the VP condition and the NP condition (all $p$s $< .001$), which indicated that the plausibility manipulation worked. The 78.3% VP attachment preference in the ambiguous condition differed from chance (both $p$s $< .001$), indicating that VP attachment was preferred to NP attachment for these items. As plausibility of the two analyses was closely matched, we conclude that our items displayed a bias toward VP attachment. This result is crucial for theories such as constraint-based theories and the unrestricted race model, which assume that on- and off-line preferences are related.

*Completion task.* An alternative means of assessing bias employed by constraint-based theorists is sentence completion (e.g., Garnsey et al. 1997; McRae et al., 1998; Trueswell et al., 1993; cf. Clifton, Kennison, & Albrecht, 1997; Pickering, Traxler, & Crocker, 2000). Eighteen participants completed booklets consisting of the 30 experimental items up to the word *with* (e.g., *The hunter killed the dangerous poacher with . . .*), together with the 81 fillers used in the eyetracking experiment, cut off at various points so that they appeared similar to the experimental items. Again, six participants saw each list, which corresponded to the lists used in the eyetracking experiment. The participants were instructed to write the first grammatical and meaningful continuation that they could think of. They were then given a second booklet containing the experimental fragments and two paraphrase fragments (e.g., *The hunter had. . .* and *The dangerous poacher had. . .*), balanced for order, and were told to indicate which of these fragments best corresponded to their completion.[1]

We excluded 2.6% of trials, on which participants said that they did not know which paraphrase was correct. Table 1 indicates how often participants employed a VP completion. This percentage was very high in all conditions and differed from chance (both $p$s $< .001$). The conditions did not differ from each other (both $p$s $>$ .1). The mean percentage of VP attachments (93.0%) was higher than in the ambiguous condition of the off-line preference task (78.3%), but, importantly, both tasks indicated a strong VP attachment preference.

[1] Note that due to the way we constructed our materials (see materials section) the NP condition in the completion was the same as the ambiguous condition for half of the items and the same as the VP condition for the other half.

*Design*

For the eye-tracking experiment, we constructed three lists of items, consisting of 10 items from each condition, with exactly one version of each item appearing in each list. Twelve participants were randomly assigned to read each list. Each list also contained 81 fillers: 45 sentences employing a range of grammatical structures and 36 items from another experiment. Thirty-four items were followed by a statement in order to encourage comprehension—10 statements followed an experimental item. Participants indicated whether the statement was a correct statement about the preceding sentence and received no feedback. Half of the statements required a "correct" response and half an "incorrect" response. The experimental items and fillers were placed in a single random order, with five fillers preceding the first experimental sentence and two fillers following the two breaks in the experiment.

*Procedure*

Participants' eye movements were recorded with a Fourward Technologies Dual Purkinje Generation 5.5 eye tracker, which monitored the right eye (though viewing was binocular). The tracker had an angular resolution of 10´ arc. A computer displayed the materials on a screen 77 cm from the participants' eyes. The screen displayed 3.8 characters per degree of visual angle. The tracker monitored participants' gaze location every millisecond and the software sampled the tracker's output to establish the sequence of eye fixations and their start and finish times.

Each participant was run individually. The experimenter told the participant to read the sentences carefully in order to understand them, but to read at his or her normal rate. Bite bars and head restraints were used to minimize head movements. Next, the participant completed a calibration procedure. Before each item or filler a calibration check was performed, and the calibration was repeated if necessary. Each sentence was presented on a single line. After reading the sentence, the participant pressed a button which led to the presentation of a comprehension statement or the next trial. The experiment, which contained two short breaks, took about half an hour.

*Analyses*

On average, the participants made 11% errors. Prior to all analyses, trials with major tracker losses and trials with fewer than four fixations were excluded. This eliminated 3% of the trials. If a fixation was shorter than 80 ms and within one character space of the previous or next fixation, it was assimilated to this fixation. All remaining fixations shorter than 80 ms were excluded. In each of the conditions, this constituted 2% of all fixations. Following Rayner and Pollatsek (1989), we assume that readers do not extract much information during such short fixations. Fixations longer than 1500 ms were also excluded. These extremely long fixations are presumably due to a major tracker loss.

We divided experimental items into seven regions, indicated by slashes in (7):

(7) The hunter/ killed/ the dangerous poacher/ with the/ rifle/ not long after/ sunset.

These regions corresponded to (1) the subject noun phrase, (2) the verb, (3) the object noun phrase, (4) the words *with the*, (5) the noun within the prepositional phrase (known as the *critical noun* region), (6) the next words (known as the *postcritical* region), and (7) the final word of the sentence. The space between words was included with the following word. Notice that the postcritical region excluded the final word of the sentence. This allowed us to interpret effects in this region as due to initial processing and hence conclude that they would not be contaminated by effects of sentence wrap-up (e.g., Just & Carpenter, 1980; Mitchell & Green, 1978; Rayner, Kambe, & Duffy, 2000; Rayner, Sereno, Morris, Schmauder, & Clifton, 1989).

We report four eye-tracking measures. *First-pass time* is the sum of all fixation times starting with the reader's first fixation inside a region until the reader's gaze leaves the region either to the left or right. For regions consisting of a single word, this corresponds to *gaze duration* (Rayner & Duffy, 1986). *First-pass regressions* include all leftward eye movements that cross the region's left boundary initiated after a first-pass fixation in the region. *Regression-path time* is the sum of all fixation times starting with the reader's first fixation inside the region until the reader's gaze leaves the region to the right (e.g., Brysbaert & Mitchell, 1996; Duffy et al., 1988; Konieczny, Hemforth, Scheepers, & Strube, 1997; Traxler, Bybee, & Pickering, 1997). For all these measures, if the eye fixated after the region before it fixated in it, the datum for this region was excluded. *Total time* is the sum of all fixations in a region. If a reader skipped a region in a measure, resulting in a zero reading time, it was excluded from the analyses for that measure. If a reader skipped two or more consecutive regions, the data of these regions and the regions following them were excluded from the analyses on that measure. We assumed that participants could not have processed the sentence completely if this was the case. The latter procedure removed less than 1% of the data for any of the analyses.[2]

Most sentence processing models do not predict in which eye-tracking measures difficulty should occur. Previous research has shown that first-pass times (e.g., Ferreira & Clifton, 1986; Frazier & Rayner, 1982; Rayner et al., 1983; Trueswell et al., 1993; 1994), first-pass regressions (e.g., Altmann et al., 1992; Altmann, van Nice, Garnham, & Henstra, 1998; Holmes & O'Regan, 1981; Rayner et al., 1983; Pickering & Traxler, 1998; Speer & Clifton, 1998), and regression-path times (e.g., Brysbaert & Mitchell, 1996; Traxler, Bybee, & Pickering, 1997; Van Gompel & Pickering, in press) can all reflect early stages of processing difficulty during syntactic ambiguity resolution, and therefore, our earliest effects may occur in any of these measures. Other models such as the competition-integration model (McRae et al., 1998; Spivey & Tanenhaus, 1998) have only modeled first-pass times, but they could, in principle, be extended to model other eye-tracking measures as well.

Using sentences comparable to (5), Traxler et al. (1998) found the earliest clear effects in the postcritical region, on the first-pass regression measure (rather than the first-pass time measure). For instance, participants experienced considerable difficulty with sentences like *The steak with the sauce that was tough didn't win a prize*, in accord with pretest data and with Gilboy et al. (1995); but this difficulty first emerged on first-pass regressions from the postcritical region after the critical word *tough* (see also Pickering & Traxler, 1998; Pickering et al., 2000, for similar effects). Therefore, we might predict effects to emerge in the postcritical region using the first-pass regression measure and the related regression-path time measure. We also predicted differences between conditions to appear on the total-time measure, for the noun region and perhaps other regions around it, in accord with Traxler et al.

### *Results and Discussion*

Table 2 presents the mean reading times and percentage regressions by condition and region. For each eye-tracking measure for each region, we conducted one-way ANOVAs with three levels of Condition (ambiguous vs VP attachment vs NP attachment), treating participants and items as random factors. Condition was treated as a within-participant and within-item variable. Table 3 presents the results of these analyses. Planned comparisons between levels are reported below.

There were no significant effects on the first-pass time measure. The pattern of first-pass regression data was more interesting. ANOVAs revealed a significant effect of condition in Region 3 (the direct object). Most likely, this is a spurious effect (note that it was before the critical region). No effects occurred in Region 4 (*with the*) or in Region 5 (the critical noun). However, significant effects emerged in Region 6 (the postcritical region). Planned comparisons revealed that, as predicted by most models, there were more regressions from the NP attachment condition than both the ambiguous condition [$F1(1, 70) = 20.1$, $p < .001$; $F2(1, 58) = 34.6$, $p < .001$] and the VP attachment condition [$F1(1, 70) = 20.9$, $p < .001$; $F2(1, 58) = 33.6$,

[2] Because of an error, the final region of one item was marginally grammatical. We excluded this complete item for the total time analyses and the final region for all other analyses.

TABLE 2

Experiment 1: Means

| | Region | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 The hunter | 2 killed | 3 the dangerous poacher | 4 with the | 5 rifle | 6 not long after | 7 sunset |
| First-pass reading times | | | | | | | |
| Ambiguous | 502 | 378 | 672 | 300 | 374 | 500 | 565 |
| VP attachment | 491 | 357 | 683 | 305 | 376 | 509 | 553 |
| NP attachment | 490 | 370 | 656 | 302 | 373 | 543 | 539 |
| First-pass regressions | | | | | | | |
| Ambiguous | | 5.0 | 14.0 | 6.0 | 8.8 | 11.3 | 42.7 |
| VP attachment | | 6.5 | 6.6 | 3.9 | 11.4 | 11.4 | 52.1 |
| NP attachment | | 7.8 | 10.0 | 3.6 | 12.2 | 25.8 | 51.2 |
| Regression-path times | | | | | | | |
| Ambiguous | 502 | 417 | 818 | 354 | 420 | 684 | 1198 |
| VP attachment | 491 | 401 | 748 | 330 | 444 | 663 | 1252 |
| NP attachment | 490 | 408 | 750 | 328 | 431 | 841 | 1491 |
| Total times | | | | | | | |
| Ambiguous | 622 | 532 | 947 | 379 | 451 | 699 | 698 |
| VP attachment | 593 | 476 | 890 | 394 | 468 | 721 | 663 |
| NP attachment | 619 | 529 | 922 | 459 | 567 | 829 | 685 |

*Note*. First pass, regression path, and total times are reported in milliseconds and first-pass regressions as the percentage of saccades leaving the region to the left after a first-pass fixation.

$p < .001$]. The ambiguous and VP conditions did not differ ($F$s $< 1$). The condition effect in the final region was also significant.[3] Planned comparisons revealed that there were fewer regressions from the ambiguous condition than the NP condition [$F1(1, 70) = 6.91$, $p = .011$; $F2(1, 56) = 5.76$, $p = .020$]). There were also fewer regressions from the ambiguous condition than the VP condition [$F1(1, 70) = 5.31$, $p = .024$; $F2(1, 56) = 4.51$, $p = .038$]. The VP and NP conditions did not differ in the final region (both $F$s $< 1$).

Regression-path times showed similar effects. In Region 3, an (unpredicted) condition effect was significant by participants, but not by items ($p = .081$). The predicted effect occurred in Region 6. Planned comparisons revealed that the NP condition was read more slowly than the ambiguous condition [$F1(1, 70) = 4.22$, $p = .044$; $F2(1, 58) = 9.12$, $p = .004$] and the VP condition [$F1(1, 70) = 7.19$, $p = .009$; $F2(1, 58) = 11.0$, $p = .002$]. The ambiguous and VP conditions did not differ ($F$s $< 1$). A similar effect appeared in the final region, with the NP condition being slower than the ambiguous condition [$F1(1, 70) = 14.1$, $p < .001$; $F2(1, 56) = 6.54$, $p = .013$] and the VP condition [$F1(1, 70) = 6.21$, $p = .015$; $F2(1, 56) = 5.73$, $p = .020$]. The ambiguous and VP conditions did not differ ($F$s $< 1$). Hence this measure showed somewhat different effects in the final region from the first-pass regressions measure.

Total-time analyses showed an effect in Region 2 (the verb), presumably due to the regressions pattern in Region 3. Clear total time effects emerged in Regions 4–6. In all these regions, the ambiguous condition was read faster than the NP condition [Region 4: $F1(1, 70) = 19.9$, $p < .001$; $F2(1, 56) = 11.22$, $p = .002$. Region 5: $F1(1, 70) = 23.4$, $p < .001$; $F2(1, 56) = 23.9$, $p < .001$. Region 6: $F1(1, 70) = 12.2$, $p < .001$; $F2(1, 56) = 16.4$, $p < .001$]. Similarly, the VP condition produced shorter times than the NP condition [Region 4: $F1(1, 70) = 12.8$, $p < .001$; $F2(1, 56) = 8.63$, $p = .005$. Region 5: $F1(1, 70) = 17.2$, $p <$

[3] The percentage of first-pass regressions from the final region was high because readers often reread the sentence after they had reached the end.

TABLE 3

Experiment 1: ANOVA Results by Region for Effect of Condition

| | $F1(2, 70)$ | $MS_e$ | $F2(2, 58)$ | $MS_e$ |
|---|---|---|---|---|
| Region 1 | | | | |
| First-pass reading times | < 1 | | < 1 | |
| Regression-path times | < 1 | | < 1 | |
| Total reading times | < 1 | | < 1 | |
| Region 2 | | | | |
| First-pass reading times | 1.20 | 5367 | 1.01 | 3324 |
| First-pass regressions | 1.35 | 62.9 | 1.05 | 60.1 |
| Regression path times | < 1 | | < 1 | |
| Total reading times | 3.38* | 12096 | 4.55* | 6907 |
| Region 3 | | | | |
| First-pass reading times | 1.33 | 4924 | < 1 | |
| First-pass regressions | 7.06** | 70.6 | 4.63* | 93.5 |
| Regression-path times | 3.69* | 14061 | 2.62 | 20901 |
| Total reading times | 2.07 | 16434 | 1.68 | 17462 |
| Region 4 | | | | |
| First-pass reading times | < 1 | | < 1 | |
| First-pass regressions | 1.38 | 70.5 | 1.46 | 33.3 |
| Regression-path times | < 1 | | 1.85 | 4079 |
| Total reading times | 11.17** | 5337 | 6.67** | 6962 |
| Region 5 | | | | |
| First-pass reading times | < 1 | | < 1 | |
| First-pass regressions | < 1 | | < 1 | |
| Regression-path times | 1.12 | 7724 | < 1 | |
| Total reading times | 13.72** | 10271 | 14.02** | 8413 |
| Region 6 | | | | |
| First-pass reading times | 1.63 | 11677 | 2.04 | 6873 |
| First-pass regressions | 13.67** | 173.4 | 22.74** | 98.3 |
| Regression-path times | 3.93* | 81795 | 6.73** | 45994 |
| Total reading times | 7.00** | 25575 | 9.61** | 14639 |
| Region 7 | | | | |
| First-pass reading times | < 1 | | < 1 | |
| First-pass regressions | 4.11* | 283.6 | 3.45* | 199.3 |
| Regression-path times | 7.30** | 117661 | 4.10* | 184031 |
| Total reading times | < 1 | | < 1 | |

*Note*. Degrees of freedom for the items analysis for Region 7 and for all regions on total time are (2, 56) (see footnote 2).
*$p$ significant at the .05 level.
**$p$ significant at the .01 level.

.001; $F2(1, 56) = 17.7$, $p < .001$. Region 6: $F1(1, 70) = 8.48$, $p = .005$; $F2(1, 56) = 12.1$, $p = .001$]. The ambiguous and VP conditions did not differ in any of these regions ($F$s $< 1$).

These results demonstrated that the NP condition became more difficult to read than either the ambiguous or the VP condition soon after the critical noun. The earliest effects of interest occurred in the postcritical region, on first-pass regressions and regression-path times, and the same effect was observed in total times for Region 4–6. These results are consistent with either a revision or a competition process. They are consistent with fixed-choice models such as the garden-path model, construal theory, and referential theory, as they predict that the VP analysis is always initially adopted. If this analysis turns out to be implausible, then processing becomes difficult. They are also consistent with variable-choice models such as the unrestricted race model. Assuming that the VP analysis is nearly always initially adopted, reanalysis should hardly ever occur in the VP condition and nearly always in the NP condition.

However, if the corpora in which NP attachment is more common than VP attachment are representative for the sentences that our participants have been exposed to, the results are incompatible with the coarse-grained frequency variant of the tuning hypothesis.

The results are consistent with constraint-based theories on the assumption that in the ambiguous condition, the VP attachment bias in combination with the plausibility of the VP analysis is strong enough to outweigh the plausibility of the NP analysis, so that there is hardly any competition.

First-pass regressions from the final region suggested that the VP condition may have been somewhat more difficult to process than the ambiguous condition, though this effect did not appear in any of the other analyses. This result is difficult for fixed-choice models to explain because the VP analysis should always be adopted and no reanalysis should be necessary in the VP condition. It also poses difficulty for constraint-based theories because they predict that, if anything, the ambiguous condition should be harder to process than the VP condition. These results fit better with variable-choice models such as the unrestricted race model, which predicts that on some occasions, readers may adopt the nonpreferred NP analysis. However, the effect was relatively weak in that it was obtained in only one analysis and at the end of the sentence. Therefore, Experiment 2 investigated balanced ambiguities. If the unrestricted race model is right, we would predict that readers adopt the NP analysis more frequently, which should result in a clear difference between the ambiguous and VP condition.

Interestingly, our earliest effects emerged in first-pass regressions and regression-path times for the postcritical region. This pattern of results is different from findings in some other eye-movement experiments, which obtained effects in first-pass times on the disambiguating region (e.g., Adams, Clifton, & Mitchell, 1998; Altmann et al., 1992, 1998; Garnsey et al., 1997; Paterson et al., 1999; Trueswell et al., 1993, 1994). One reason may be that we employed semantic rather than syntactic disambiguation, which was employed in many other studies. Semantic information may become available more slowly than syntactic information, and it is certainly less strong in that it usually does not fully disambiguate. Furthermore, although plausibility was manipulated on the noun in the PP, the actual disambiguation may have been at the next word. For example, (4b) could still be plausibly continued as *The hunter killed the leopard with the rifle wound*. This might explain why our earliest effects were obtained in the postcritical region. The fact that our results occurred in first-pass regressions and regression-path times from this region suggests that readers went back to the parts of the sentence that caused the disruption.

However, it does not explain why a number of other studies on VP–NP attachment ambiguities found effects in first-pass on the disambiguating noun (Britt et al., 1992; Ferreira & Clifton, 1986; Ni et al., 1996; Rayner et al., 1983). We think there are two reasons. In some studies (Britt et al., 1992; Rayner et al., 1983), the disambiguating noun occurred at the end of a clause. Reading times at the end of a clause are known to include more integrative “wrap-up” processing (Just & Carpenter, 1980; Mitchell & Green, 1978; Rayner et al., 1989, 2000) and therefore reflect later processing. Furthermore, the critical regions in most studies on VP–NP ambiguities (and indeed most studies on other syntactic ambiguities) were much longer than in the current experiment. For instance, Rayner et al.’s (1983) first-pass effects were based on a critical region including the following *clause*. Therefore, they include much later processing than the first-pass times in our experiment. Similarly, Ni et al. (1996) included words preceding the critical phrase, and their phrases sometimes contained adjectives that may have disambiguated the structure before the critical noun. Finally, Ferreira and Clifton’s (1986) critical region employed syntactic disambiguation and consisted of at least three words. When one uses a short disambiguating region consisting of only one word as in our experiment, it is in fact not too surprising to find the effects in the following region. First-pass reading times for one-word regions are often very short, and therefore readers will often have planned a saccade before the processor has evaluated the plausibility of the syntactic analysis. This explanation fits well with current models of eye-movement control such as the E-Z reader model (Reichle, Pollatsek, Fisher, & Rayner, 1998), which claims that saccades to the next word are programmed before lexical access is complete when accessing the word is slow (e.g., because the word is infrequent). Assuming that syntactic and semantic processing follow lexical access, this predicts that the evaluation of the plausibility of a syntactic structure may also lag behind the programming of saccades.

Indeed, a number of eye-movement experiments that employed a short critical region in combination with semantic disambiguation also showed the earliest effects in regression-based measures for the postcritical region (Pickering et al., 2000; Pickering & Traxler, 1998, Traxler et al., 1998). The results of Experiment 1 provide an important indication of the way in which any effects might reveal themselves in the theoretically more critical sentences used in Experiment 2.

## EXPERIMENT 2

This eye-tracking experiment investigated VP–NP attachment ambiguities in which there was no initial preference for either VP or NP attachment. It employed items that were closely related to those in Experiment 1 but included the additional word *only*. For these sentences, current models of syntactic ambiguity resolution make very different predictions about which sentences would cause processing difficulty.

### *Participants*

Twenty-seven participants took part. Six additional participants were excluded from the analyses, three because their data reflected a high percentage of tracker losses and three because they answered more than 25% of the statements incorrectly.

### *Items*

Participants read 30 sentences like (5), repeated below (see Appendix):

(5a) The hunter killed only the poacher with the rifle not long after sunset. (ambiguous)

(5b) The hunter killed only the leopard with the rifle not long after sunset. (VP attachment)

(5c) The hunter killed only the leopard with the scars not long after sunset. (NP attachment)

The items were semantically disambiguated to the VP or NP analysis or were semantically ambiguous (see plausibility pretest). The items were similar to those in Experiment 1, except that the object noun phrase began with the word *only* and did not include an adjective. The word-form frequency of the noun in the PP did not differ according to the Celex database [$F(1, 29) = 1.15$, $p = .29$, $MS_e = 110039$], although there was a difference in the logarithm of the word-form frequency [$F(1, 29) = 5.20$, $p = .03$, $MS_e = .63$]. The fact that the noun was somewhat more frequent in the NP condition than in the ambiguous and VP condition works against the prediction of the unrestricted race model, which claims that the NP condition should be more difficult than the ambiguous condition.

### *Pretests*

Pretests were the same as in Experiment 1, except as follows.

*Plausibility pretest.* Eighteen participants rated the plausibility of 50 sets of items. For the VP analysis, we replaced *with* with *by using* or with *when using* (4 item sets). For the final 30 sets of items, Table 4 shows the mean plausibility rating for each condition, for each analysis.

We again conducted 2 (Paraphrase type: VP–PP vs NP–PP paraphrase) × 3 (Condition: ambiguous vs VP vs NP) ANOVAs and found a significant interaction [$F1(2, 34) = 99.7$, $p < .001$, $MS_e = 1.12$; $F2(2, 58) = 714$, $p < .001$, $MS_e = .26$]. Planned comparisons revealed plausibility differences between the two paraphrase conditions for VP attachment and for NP attachment (all $p$s $< .001$). In the ambiguous conditions, the numerical difference between the paraphrases was very small, although the ef-

TABLE 4

Experiment 2: Item Characteristics

| | Condition | | |
|---|---|---|---|
| | Ambiguous | VP attachment | NP attachment |
| Plausibility pretest | | | |
| Plausibility VP-PP paraphrase | 5.8 | 5.8 | 2.2 |
| Plausibility NP-PP paraphrase | 5.3 | 2.5 | 5.9 |
| Frequency of noun in PP | 185 | 185 | 277 |
| Log frequency of noun in PP | 1.72 | 1.72 | 2.18 |
| Off-line preference task | | | |
| Percentage of VP attachment | 55.0 | 94.4 | 2.8 |
| Completion task | | | |
| Percentage of VP completions | 37.2 | 46.1 | 37.4 |

*Note*. Frequency is the number of occurrences of the word form in the Celex database (17.9 million words).

fect was significant by items [$F1(1, 34) = 1.81$, $p = .19$; $F2(1, 58) = 12.9$, $p < .001$].

We analyzed the absolute differences between the two paraphrases in the three attachment conditions. Planned comparisons revealed that the difference between the paraphrases was much smaller in the ambiguous condition than in the VP condition and the NP condition (all $p$s $< .001$).

*Off-line preference task*. This pretest employed 18 further participants. The mean percentages of VP–PP paraphrases are presented in Table 4.[4] ANOVAs revealed a main effect of condition [$F1(2, 34) = 322$, $p < .001$, $MS_e = 118$; $F2(2, 58) = 174$, $p < .001$, $MS_e = 364$]. Planned comparisons revealed that the percentage of VP–PP attachment responses in the ambiguous condition differed from those in both the VP attachment condition and the NP attachment condition (all $p$s $< .001$), which indicated that the plausibility manipulation worked.

Importantly, the 55.0% VP–PP paraphrase choice in the ambiguous condition did not differ from chance (both $F$s $< 1.2$). This indicates that plausibility and preference information did not favor either the VP or NP analysis. Additionally, participants hardly ever took the implausible analysis in the VP and NP conditions: 5.6 and 2.8% respectively. We concluded that there was no preference for either VP or NP attachment in the ambiguous condition for Experiment 2, but that the final analysis in the disambiguated conditions was nearly always determined by plausibility.

*Completion task*. Eighteen participants took part. We excluded 0.2% of trials, on which participants said that they did not know which paraphrase was correct. The overall number of VP completions was significantly less than chance on the item analysis, but not on the participant analysis [$F1 = 1.62$, $p = .22$, $MS_e = 527$; $F2 = 11.44$, $p = .002$, $MS_e = 124$] (see Table 4). There were no significant differences between the conditions (both $p$s $> .10$). The mean percentage of VP completions, 40.3%, is within the range normally considered balanced (e.g., Rayner & Duffy, 1986). We therefore concluded that there was no clear preference for either VP or NP attachment.

### *Predictions*

The off-line preference and completion task indicated that there was no bias for either VP or NP attachment for these items. The plausibility pretest and off-line preference task showed that plausibility favored one analysis over the other in the VP and NP conditions and that it equally favored both analyses in the ambiguous condition. Hence, constraint-based the-

[4] One additional participant was excluded because she chose the NP-PP paraphrase 93% of the time, a pattern clearly different from any other participant. We presumed that she had failed to understand the task. Including this participant would have taken the percentage of VP completions closer to 50%.

ories predict competition when the items are semantically ambiguous [as in (5a)], but much less competition when the items are disambiguated [as in (5b) and (5c)]; thus (5a) should cause difficulty compared with (5b) and (5c), which should not differ. Garden-path theory and construal predict that the parser initially adopts the VP analysis and only reanalyzes if that analysis becomes implausible. Hence, they predict difficulty with the NP condition (5c) versus the other two conditions [(5a) and (5b)], which should not differ. Ni et al.'s (1996) version of referential theory predicts that the parser initially adopts the NP analysis and only reanalyzes if that analysis becomes implausible. Hence, it predicts difficulty with the VP condition (5b) versus the other two conditions [(5a) and (5c)], which should not differ. The unrestricted race model predicts that each analysis is initially adopted about half the time because there is no initial bias for either analysis. The initial analysis is implausible half the time in the disambiguated conditions, but always plausible in the ambiguous condition. Hence, processing difficulty will occur in the disambiguated conditions (due to reanalysis), but not in the ambiguous condition. Assuming that Experiment 1 demonstrated the effects of difficulty on patterns of eye movements for VP/NP ambiguities, we would expect any differences among the conditions to begin to emerge in first-pass regressions and regression-path time for the postcritical region.

### *Design, Procedure, and Analyses*

These were the same as Experiment 1, except as follows. The word *only* was included in the direct object region. Nine participants were randomly assigned to read each list. There were 66 filler items. Thirty items, including 10 experimental items, were followed by a comprehension statement with feedback. There was only one programmed break in the experiment. Six fillers preceded the first experimental sentence and three followed the break. On average, participants made 8% errors. In each condition, 4% of fixations was excluded because their duration was shorter than 80 ms. One percent of the trials were excluded because they had fewer than four fixations or major tracker losses. Less than 1% of the data was removed as a result of two consecutive regions having zero reading times.

### *Results and Discussion*

Table 5 presents the mean reading times and percentage regressions by condition and region. For each eye-tracking measure for each region, we conducted one-way ANOVAs with three levels of Condition (ambiguous vs VP attachment vs NP attachment). Table 6 presents the results of these analyses. Planned comparisons between levels are reported below.

As in Experiment 1, first-pass times did not produce any significant effects. However, the first-pass regression measure showed clear effects in Region 6 (the postcritical region), with planned comparisons showing that fewer regressions occurred in the ambiguous condition than in both the VP condition [$F1(1, 52) = 4.27$, $p = .044$; $F2(1, 58) = 6.23$, $p = .015$] and the NP condition [$F1(1, 52) = 12.1$; $p < .001$; $F2(1, 58) = 13.4$, $p < .001$]. The VP and NP conditions did not differ [$F1(1, 52) = 1.99$, $p = .16$; $F2(1, 58) = 1.36$, $p = .25$].

The pattern for regression-path times was very similar. In Region 6, the condition effect was significant by items and just missed significance by participants ($p = .071$). Planned comparisons showed that the ambiguous condition was faster than both the VP condition [$F1(1, 52) = 4.42$, $p = .041$; $F(1, 58) = 10.0$, $p = .002$] and the NP condition [$F1(1, 52) = 3.91$, $p = .053$; $F(1, 58) = 9.04$, $p = .004$]. The VP and NP conditions did not differ ($F$s $< 1$).

Total times showed an effect in Region 4 (*with the*). Planned comparisons showed that the ambiguous condition was read faster than the VP condition [$F1(1, 52) = 6.24$, $p = .016$; $F2(1, 58) = 5.70$, $p = .020$] and the NP condition [$F1(1, 52) = 5.91$, $p = .019$; $F2(1, 58) = 5.40$, $p = .024$], which did not differ from each other ($F$s $< 1$). There was a marginally significant effect in Region 3, the direct object ($p < .10$ by participants and items); the means suggest that the ambiguous condition may have been read faster than the disambiguated conditions. No other effects approached significance by participants and items.

TABLE 5

Experiment 2: Means

| | Region | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 The hunter | 2 killed | 3 only the poacher | 4 with the | 5 rifle | 6 not long after | 7 sunset |
| First-pass reading times | | | | | | | |
| Ambiguous | 371 | 331 | 520 | 265 | 310 | 448 | 464 |
| VP attachment | 398 | 315 | 555 | 272 | 317 | 463 | 479 |
| NP attachment | 405 | 312 | 560 | 286 | 314 | 442 | 491 |
| First-pass regressions | | | | | | | |
| Ambiguous | | 6.2 | 9.0 | 6.8 | 15.6 | 13.9 | 59.1 |
| VP attachment | | 9.8 | 9.0 | 7.4 | 17.1 | 22.1 | 56.4 |
| NP attachment | | 6.5 | 9.5 | 9.1 | 14.3 | 27.1 | 61.1 |
| Regression-path times | | | | | | | |
| Ambiguous | 371 | 363 | 620 | 299 | 360 | 562 | 1063 |
| VP attachment | 398 | 374 | 626 | 332 | 379 | 693 | 1172 |
| NP attachment | 405 | 342 | 645 | 342 | 377 | 682 | 1133 |
| Total times | | | | | | | |
| Ambiguous | 421 | 425 | 706 | 362 | 398 | 633 | 595 |
| VP attachment | 469 | 435 | 795 | 409 | 404 | 681 | 597 |
| NP attachment | 456 | 412 | 764 | 410 | 432 | 674 | 611 |

*Note*. First pass, regression path, and total times are reported in milliseconds and first-pass regressions as the percentage of saccades leaving the region to the left after a first pass fixation.

The data revealed that the ambiguous condition was easier to process than both the VP and NP attachment conditions. There was no evidence for either a VP or NP attachment preference: Both conditions were read equally fast. The earliest effects arose in first-pass regressions and regression-path times immediately after the critical noun. Thus the locus of the effects was identical to that in Experiment 1.

The results fit well with a variable-choice model such as the unrestricted race model (Traxler et al., 1998; Van Gompel et al., 2000). In this model, readers may adopt either the VP or NP analysis initially. If the initial analysis turns out to be inconsistent with later information, reanalysis occurs and processing difficulty ensues. Given that there was no difference between the VP and NP condition, our results suggest that reanalysis occurred equally often in the VP and NP condition and therefore that the VP and NP analysis were adopted about equally often during the first stage of processing. This is consistent with the off-line and completion pretests, which indicated that the materials from Experiment 2 were balanced.

The results may also be compatible with other variable-choice models. For example, some variants of the tuning hypothesis may claim that readers are sensitive to fine-grained frequency information associated with *only* and that VP and NP attachment in sentences containing *only* are about equally frequent in English. Hence, reanalysis should occur about equally often in the VP and NP disambiguated conditions. However, the results are more difficult to reconcile with the coarse-grained frequency variant of the tuning hypothesis that was favored by Mitchell et al. (1995; Brysbaert & Mitchell, 1996; Cuetos et al., 1996). According to this variant, the processor ignores lexical preferences and preferences associated with particular NPs. Thus, VP/NP attachment ambiguities should always show the same preference, regardless of whether the direct object NP contained an adjective (Experiment 1) or *only* (Experiment 2). However, Experiment 1 showed a strong VP preference, whereas there was no preference in Experiment 2. In order to save this variant of the tuning hypothesis, one would have to assume that

TABLE 6

Experiment 2: ANOVA Results by Region for Effect of Condition

| | $F1(2, 52)$ | $MS_e$ | $F2(2, 58)$ | $MS_e$ |
|---|---|---|---|---|
| Region 1 | | | | |
| First-pass reading times | 2.47 | 2769 | 1.91 | 4664 |
| Regression-path times | 2.47 | 2769 | 1.91 | 4664 |
| Total reading times | 2.47 | 5451 | 1.46 | 10720 |
| Region 2 | | | | |
| First-pass reading times | < 1 | | 1.87 | 1977 |
| First-pass regressions | 2.29 | 53.1 | 1.52 | 78.9 |
| Regression-path times | 1.82 | 3528 | 2.14 | 3999 |
| Total reading times | < 1 | | < 1 | |
| Region 3 | | | | |
| First-pass reading times | 2.95 | 4006 | 1.69 | 7769 |
| First-pass regressions | < 1 | | < 1 | |
| Regression-path times | < 1 | | < 1 | |
| Total reading times | 3.10 | 16752 | 2.72 | 21394 |
| Region 4 | | | | |
| First-pass reading times | 2.30 | 1035 | 1.74 | 1706 |
| First-pass regressions | < 1 | | < 1 | |
| Regression-path times | 1.54 | 8908 | 1.20 | 13745 |
| Total reading times | 4.05* | 5112 | 3.70* | 6567 |
| Region 5 | | | | |
| First-pass reading times | < 1 | | < 1 | |
| First-pass regressions | < 1 | | < 1 | |
| Regression-path times | < 1 | | < 1 | |
| Total reading times | 1.69 | 4458 | 1.85 | 6743 |
| Region 6 | | | | |
| First-pass reading times | < 1 | | < 1 | |
| First-pass regressions | 6.12** | 181.1 | 7.00** | 200.2 |
| Regression-path times | 2.78 | 45505 | 6.37** | 27139 |
| Total reading times | < 1 | | 1.88 | 10018 |
| Region 7 | | | | |
| First-pass reading times | < 1 | | < 1 | |
| First-pass regressions | < 1 | | < 1 | |
| Regression-path times | < 1 | | < 1 | |
| Total reading times | < 1 | | < 1 | |

* *p* significant at the .05 level.
** *p* significant at the .01 level.

the participants from Experiment 1 and 2 had different preferences, but this seems unlikely because they were randomly drawn from the same population. In addition, the results from our pretests, which used different participants, were consistent with the eye-movement experiments.

The results are also inconsistent with fixed-choice models, as such models predict that the processor always adopts the same analysis, so that difficulty relative to the ambiguous condition occurs either for the VP or NP condition, but not for both. The data are incompatible with the garden-path theory, which predicts that the processor adopts the VP analysis initially and only reanalyzes when this analysis is implausible. Thus, the garden-path model cannot explain why the VP attachment condition is more difficult than the ambiguous condition. Similarly, the results are incompatible with referential theory, as interpreted by Ni et al. (1996), which predicts that the processor adopts the NP analysis initially.

Importantly, the current results provide evidence against fixed-choice models in an ambiguity that involves primary phrases. Traxler et al. (1998) showed that in ambiguities that involved nonprimary relations, globally ambiguous sentences were easier to read than disambiguated sentences (when there was no strong preference for either analysis). However, it was unclear whether this kind of pattern could also be obtained for ambiguities that involved primary relations. Construal theory (Frazier & Clifton, 1996) predicts a different pattern for these two types of ambiguities because it is a variable-choice model with respect to ambiguities that involve only nonprimary relations, but a fixed-choice model with respect to ambiguities that involve at least one primary relation. It claims that the VP/NP attachment ambiguities in the current experiments involve two primary relations and that minimal attachment should apply. However, the fact that the VP condition was more difficult than the ambiguous condition indicates that our readers initially adopted the NP analysis on a number of trials. Hence, the current experiment suggests that nonprimary ambiguities do not constitute a special class: Even in ambiguities involving primary relations, globally ambiguous sentences are easier to process than disambiguated sentences.

Constraint-based theories (e.g., MacDonald et al., 1994; McRae et al., 1998; Spivey & Tanenhaus, 1998; Spivey-Knowlton & Sedivy, 1995; Tabor et al., 1997; Tabor & Tanenhaus, 1999) also cannot satisfactorily account for the data. Such models predict that the ambiguous condition should cause processing difficulty because there is no initial preference for either attachment site, nor does plausibility favor one analysis over the other. This ought to result in competition between the VP and NP analyses. In contrast, the VP and NP analyses should be relatively easy because plausibility supports only one analysis. But in fact the exact opposite occurred: The ambiguous condition was the easiest.

In theory, there are a number of ways that these results might still be consistent with competition models. In the introduction, we mentioned that some constraint-based theories allow competition to decrease very rapidly (e.g., McRae et al., 1998; Spivey & Tanenhaus, 1998; Tabor & Tanenhaus, 1999). These models might claim that competition occurred as soon as the ambiguity arose at *with*, but that one analysis became more activated than its alternative during the next words, so that competition was relatively weak when readers reached the critical noun *rifle* in the ambiguous condition. Such models predict that, especially if items are not well balanced, the preferred analysis should quickly become more activated than the nonpreferred analysis. Thus, less well-balanced items should produce weaker competition than more closely balanced items.

In order to investigate this, we split the materials into balanced and biased items. The bias of the materials was determined on the basis of the results from the eye-tracking study itself. Because first-pass regressions from the postcritical Region 6 was the earliest measure to show significant effects, balanced materials were defined as those 15 materials for which the difference between the VP and NP attachment in first-pass regressions from Region 6 was smallest (a difference of 12.5% or less). The other 15 materials were assigned to the biased group. Table 7 presents the means for each condition for each group of materials.

ANOVAs with Condition as a within-participants and -items factor and Item Bias (balanced vs biased) as a within-participants and between-items factor were conducted. They revealed an effect of Item Bias, significant by participants only [$F1(1, 26) = 8.44$, $p = .007$, $MS_e = 309$; but $F2(1, 28) = 1.90$, $p = .18$, $MS_e = 423$]. If this is a real effect, it could be due to irrelevant differences between the two groups of items (e.g., frequency or length). They also revealed an effect of Condition [$F1(2, 52) = 7.31$, $p = .002$, $MS_e = 389$; $F(2, 56) = 6.90$, $p = .002$, $MS_e = 203$], just as in the original analyses. There was no interaction between the two factors [$F1(2, 52) = 2.39$, $p = .10$, $MS_e = 290$; $F2(2, 56) = .60$, $p = .55$, $MSe = 203$], which indicates that the pattern of results was roughly similar for balanced and biased items.[5] Most impor-

[5] The NP attachment condition seemed to produce more regressions than the VP attachment condition in the biased materials, but not in the balanced materials. This was due to the fact that, overall, there was a slight (nonsignificant) preference for VP attachment. As a consequence of removing all balanced materials, the difference increased for the biased materials.

TABLE 7

Experiment 2: Percentages of First-Pass Regressions from Region 6 by Condition and Item and Participant Bias

| | Item bias | | Participant bias | |
|---|---|---|---|---|
| | Balanced | Biased | Balanced | Biased |
| Ambiguous | 11.2 | 16.5 | 14.8 | 13.1 |
| VP attachment | 21.4 | 22.8 | 24.0 | 20.2 |
| NP attachment | 21.5 | 32.8 | 26.2 | 27.9 |

tantly, when the balanced items were analyzed separately, the effect of Condition was still significant [$F1(2, 52) = 3.52$, $p = .037$, $MS_e = 302$; $F2(2, 28) = 6.99$, $p = .003$, $MS_e = 80.4$] Thus, even for the balanced items the ambiguous condition produced fewer regressions than either the VP or NP attachment condition. In fact, the pattern for the balanced items was the same to that for the complete set of items. The results do not show any evidence that competition is stronger for more balanced items.

Just as items can be biased, so participants may have a strong preference for one of the analyses. Hence, we also tested whether participants that did not have a strong preference for one analysis experienced more competition than participants that did have a strong preference. Participants were split into two groups. Bal-anced participants were defined as those 13 participants for which the difference between the VP and NP attachment in first-pass regressions from Region 6 was smallest (12.2% or less). Biased participants were the 14 other participants. The means for each group of participants are presented in Table 7. ANOVAs with Condition as a within-participants and -items factor and Participant Bias (balanced vs biased) as a between- and within-items factor showed an effect of Condition [$F1(2, 50) = 5.98$, $p = .005$, $MS_e = 186$, $F2(2, 58) = 6.15$, $p = .004$, $MS_e = 383$], but no effect of Participant Bias [$F1(1, 25) = .050$, $p = .83$, $MS_e = 676$; $F1(1, 29) = 1.41$, $p = .24$, $MS_e = 208$] or interaction between the two factors [$F(2, 50) = .37$, $p = .69$, $MS_e = 186$; $F2(2, 58) = .53$, $p = .59$, $MS_e = 383$], indicating that biased and balanced participants performed similarly. As shown in Table 7, the pattern for the balanced participants was very similar to the overall pattern: The ambiguous condition produced fewer regressions than both the VP and NP attachment conditions. Separate analyses of the balanced participants showed an effect of Condition [$F1(2, 58) = 3.13$, $p = .051$, $MS_e = 294$, $F2(2, 24) = 4.97$ $p = .016$, $MS_e = 97.4$]. Thus, there was no evidence for competition, even for the most balanced participants.

Finally, we conducted a more direct test of models that claim that competition decreases quickly. If competition decreases rapidly, the preferred analysis should quickly become more activated than its alternative, and therefore the activation of the preferred analysis should be much higher than the activation of the nonpreferred analysis just before the critical word *rifle*/*scars*. Hence, disambiguation toward the preferred analysis should be easier than disambiguation toward the nonpreferred analysis. In addition, disambiguation toward the preferred analysis should not be more difficult than a globally ambiguous sentence, where there should be some (though perhaps weak) competition due to equal plausibility of the two analyses. We tested this by recoding the disambiguated conditions into cases where the disambiguation was consistent with the bias and into cases where the disambiguation was inconsistent with it. Thus, in the consistent condition, the sentence was either VP biased and VP disambiguated, or NP biased and NP disambiguated. In the inconsistent condition, the disambiguation was different from the bias. The completions were used to determine the bias of the materials. As the huge difference in percentage of VP completions between Experiment 1 and 2 shows, this task is sensitive to item biases. Furthermore, constraint-based theories claim that on-line biases should reflect completion data (e.g., Garnsey et al., 1997; McRae et al., 1998; Trueswell et al., 1993).[6]

[6] It is likely that the on-line data provide a better estimate of the item biases. However, we could not use them here, because the ambiguous-disambiguated difference and VP-NP difference are not independent. For example, selecting those materials where the VP condition is easier than the NP condition necessarily reduces the VP-ambiguous difference. Hence, even if the item biases are caused by random variability, this would still reduce the difference between the ambiguous and consistent condition.

Three items were excluded from the analyses, because they were perfectly balanced. As before, we analyzed first-pass regressions from the postcritical region. The mean percentages of regressions for the ambiguous, consistent, and inconsistent conditions were 14.5, 26.3, and 23.9% respectively. Analyses with condition (ambiguous vs consistent vs inconsistent) as a within-participants and -items factor showed a main effect of condition: [$F1(2, 52) = 4.76$, $p = .013$, $MS_e = 168$; $F2(2, 52) = 5.23$, $p = .008$, $MS_e = 223$]. However, planned comparisons showed that the two disambiguated conditions did not differ ($F$s $< 1$). This suggests that the materials were very homogeneous just before the critical noun—no or very few materials were strongly biased at this point. Further comparisons showed that the ambiguous condition produced fewer regressions than both the inconsistent condition [$F1(1, 52) = 5.78$, $p = .020$; $F2(1, 52) = 5.95$, $p = .018$] and the consistent condition [$F1(1, 52) = 8.28$, $p = .006$; $F2(1, 52) = 9.35$, $p = .004$]. These results challenge constraint-based theories in which competition is resolved rapidly because they predict that, if anything, the ambiguous condition should be more difficult than the consistent condition. No competition should occur in the consistent condition, because the initial bias and plausibility favored the same analysis. In contrast, some competition should have occurred in the ambiguous condition, because plausibility supported both analyses. (Note that the nonpreferred analysis should not have been completely abandoned according to current constraint-based theories; see the introduction). Clearly, the data did not support this prediction.

Finally, we need to consider whether the results might be due to a frequency difference between the NP attachment and ambiguous conditions for the noun in the PP. Although the raw frequency of this noun was controlled for both conditions, there was a difference in the logarithmic frequency: The logarithmic frequency was higher in the NP than in the ambiguous condition. If this small difference had an effect, we would expect the NP condition to be easier than the ambiguous condition, but this was not the case. The noun frequency of the NP condition was also higher than the VP condition. Possibly, this might have neutralized a difference between the two conditions. However, it is important to note that the crucial contrast was between the VP and ambiguous condition. In these conditions, the noun was identical.

## GENERAL DISCUSSION

Experiment 1 used VP–NP attachment ambiguities with an initial bias for VP attachment. It found that sentences that were semantically disambiguated to the NP analysis rapidly became harder to process than sentences that were disambiguated to the VP analysis or ambiguous sentences in which either analysis was plausible. Experiment 2 used similar ambiguities, but in this case, there was no bias for either VP or NP attachment. It found that ambiguous sentences were easier to process than either type of disambiguated sentence (which did not differ in difficulty). The effects in Experiment 2 were obtained in the same eye-movement measures and regions as in Experiment 1, indicating that syntactic ambiguity resolution in balanced ambiguities is roughly as rapid as in biased ambiguities.

We have argued that the results from Experiment 2 provide evidence against competition. Hence, the results are incompatible with constraint-based theories (e.g., MacDonald et al., 1994; McRae et al., 1998; Spivey & Tanenhaus, 1998; Spivey-Knowlton & Sedivy, 1995; Tabor et al., 1997; Tabor & Tanenhaus, 1999) because a fundamental assumption of these theories is that the processor employs competition as a mechanism of syntactic ambiguity resolution. Given that it is such a crucial assumption, it is hard to see how such theories can account for our results. For example, without competition, the competition-integration model that was proposed by McRae et al. (1998) and Spivey and Tanenhaus (1998) would have to become an entirely different type of model; in fact, it is unclear how the model would work without it. Although it may be turn out to be possible to develop alternative parallel interactive theories without a competition mechanism (e.g., Pearlmutter & Mendelsohn, 1999, for an initial attempt), these theories are very different from the current constraint-based theories. Further development of these models is required before it can be established whether they can account for our data.

Another problem for constraint-based theories is that our results suggest that syntactic ambiguity resolution cannot be equated with lexical ambiguity resolution, contrary to their claims (e.g., MacDonald et al., 1994; Trueswell et al., 1993, 1994). Research on lexical ambiguity resolution suggests that balanced words in neutral contexts are harder to process than balanced words in constraining contexts or biased words (Duffy et al., 1988; Rayner & Duffy, 1986). In lexical ambiguity resolution, it appears likely that the lack of evidence in favor of one or other interpretation causes competition between the meanings of the word and hence processing difficulty. However, the ambiguous condition in Experiment 2 reproduced these conditions in syntactic ambiguity resolution and found very different results; there was no evidence for competition in truly ambiguous sentences compared to semantically disambiguated sentences. In fact, the ambiguous condition of Experiment 2 was easier than both disambiguated conditions. We conclude that lexical and syntactic ambiguity resolution are very different processes.

Our results are also problematic for fixed-choice two-stage models such as the garden-path theory (e.g., Frazier, 1979, 1987) and referential theory (Crain & Steedman, 1985; Altmann & Steedman, 1988; Ni et al., 1996). Such models predict that readers should always initially adopt the same analysis. Although this prediction is consistent with the findings from Experiment 1, which showed little difference between the ambiguous and VP condition, it is not consistent with the findings of Experiment 2: Both the VP and NP conditions were more difficult to read than the ambiguous condition. The VP–NP attachment ambiguity showed that results for primary relations are very similar to those obtained by Traxler et al. (1998) for nonprimary relations. For both types of ambiguities, strongly biased globally ambiguous sentences are easier than biased sentences that are disambiguated toward their nonpreferred analysis, but about as easy as biased sentences that are disambiguated toward their preferred analysis. This contrasts with ambiguities that are roughly balanced. In such ambiguities, globally ambiguous sentences are more difficult than disambiguated sentences, whichever way they are disambiguated. This finding is inconsistent with construal theory (Frazier & Clifton, 1996), which claims that the processor is variable choice for ambiguities involving nonprimary relations, but fixed choice for ambiguities involving primary relations.

The results fit better with variable-choice models, which claim that the processor does not adopt the same analysis on all trials. However, they are not compatible with all variable-choice models. The results are difficult to reconcile with the coarse-grained frequency variant of the tuning hypothesis, as it incorrectly predicts that the sentences in Experiment 1 and 2 should show the same preferences, because fine-grained frequency information such as the presence or absence of *only* in a NP is not used in initial parsing.

We have proposed a particular variant of variable-choice models, the unrestricted race model (see also Traxler et al., 1998; Van Gompel et al., 2000). This model fits well with the results from the current experiments and the experiments on relative clause ambiguities in Traxler et al. (1998). The unrestricted race model claims that the analyses of a syntactic ambiguity are engaged in a race and that the analysis that is constructed fastest is adopted. How often a particular analysis is adopted depends on the strength of the initial bias. If there is no clear initial bias for either analysis, as in Experiment 2, each analysis will be adopted about half the time. When the sentence is semantically disambiguated toward either VP or NP attachment, this results in reanalysis about half the time. In contrast, when there is an initial bias for one analysis, as in Experiment 1, the preferred analysis is usually constructed fastest. Hence little or no processing difficulty arises when the ambiguity is disambiguated toward this analysis. In general, the stronger the ambiguity is biased, the more often the favored analysis is initially selected, the more often reanalysis is required when the sentence is disambiguated toward the disfavored analysis, and the less often reanalysis is required when it is disambiguated toward the favored analysis. Thus, the unrestricted race model claims that graded effects in syntactic ambiguity resolution are at least in part due to the frequency with which the initial analysis has to be revised.

In the current study, the off-line and completion pretests were consistent with on-line preferences as measured by the eye-tracking experiments. However, it is important to stress that our conclusions do not depend on the pretests. The eye-tracking experiments themselves showed that the materials in Experiment 1 were biased (because the VP and NP conditions differed) and the materials in Experiment 2 balanced (because the VP and NP conditions did not differ). Most importantly, they showed that the ambiguous condition was easier than both disambiguated conditions in balanced ambiguities. It is these eye-tracking results which provide the most critical evidence against competition and fixed-choice models, not the correspondence between our on-line data and pretests.

We favor the unrestricted race model over an account which assumes that readers were just lazy and did not resolve the ambiguous conditions in the experiments. We think that there are a number of reasons why such an explanation is not very attractive. Note that it would be inconsistent with all current sentence processing models. All two-stage and constraint-based theories assume that the same processing mechanisms deal with globally ambiguous and disambiguated sentences. One would have to postulate an additional mechanism which explains why globally ambiguous sentences are not resolved, whereas disambiguated sentences are. One would also need a mechanism that determines when a sentence is sufficiently ambiguous that processing be suspended. Hence, such a model would clearly be less parsimonious than models that have a single mechanism to handle both types of sentences. We also think that this interpretation of our data is unlikely. First, we know that readers did attempt to resolve the disambiguated conditions, because they showed processing difficulty relative to the ambiguous condition. Second, recognizing that a sentence is globally ambiguous would require constructing all relevant analyses, the VP and NP analyses in the current case. The processor cannot simply have failed to analyze the ambiguous sentences at all. Finally, this explanation would predict that suspension of processing also takes place in globally ambiguous sentences that are implausible on either analysis. Thus, sentences such as (8) should be easy to process as well:

(8) The hunter killed only the leopard with the rainbow.

Although we did not investigate sentences that were implausible on both readings, we think this prediction is very unlikely to be correct and that data from sentences such as (8) would support the unrestricted race model, which claims that they should be hard.

To summarize, our results provide evidence about the mechanisms that underlie parsing. In particular, the finding of Experiment 2 that balanced ambiguous VP–NP sentences are easier to process than their disambiguated controls provides clear evidence against both the mechanism of competition assumed by constraint-based models and the strategies assumed by fixed-choice two-stage models. In contrast, we have argued that the processing is consistent with the unrestricted race model, a variable-choice two-stage model of parsing.

## APPENDIX

*Experimental Items from Experiment 1*

The farmer injured the young thief with the staff only a few days ago.

The farmer injured the black sheep with the staff only a few days ago.

The farmer injured the black sheep with the wound only a few days ago.

The outlaw wounded the tall guard with the crossbow during the night.

The outlaw wounded the grey horse with the crossbow during the night.
The outlaw wounded the tall guard with the whiskers during the night.

The knight struck the hated king with the sword during the dark night.
The knight struck the iron hinge with the sword during the dark night.
The knight struck the hated king with the beard during the dark night.

The explorer found the lost tent with the charts after a long search.
The explorer found the large sea with the charts after a long search.
The explorer found the large sea with the whales after a long search.

Patrick transported the black car with the trailer the day before yesterday.
Patrick transported the black cow with the trailer the day before yesterday.
Patrick transported the black car with the scratch the day before yesterday.

The policeman touched the dirty beggar with the stick after some hesitation.
The policeman touched the dirty stains with the stick after some hesitation.
The policeman touched the dirty beggar with the fever after some hesitation.

The hooligan damaged the new shop with the fireworks late at night.
The hooligan damaged the new pane with the fireworks late at night.
The hooligan damaged the new shop with the discounts late at night.

The farm-hand flogged the wicked boy with the branch a number of times.
The farm-hand flogged the young pony with the branch a number of times.
The farm-hand flogged the young pony with the bruise a number of times.

The handyman repaired the enormous toolbox with the pliers right after lunch.
The handyman repaired the electric circuit with the pliers right after lunch.
The handyman repaired the electric circuit with the damage right after lunch.

The walker saw the spying man with the telescope early in the afternoon.
The walker saw the young deer with the telescope early in the afternoon.
The walker saw the young deer with the infection early in the afternoon.

The farmer deterred the cunning robber with the revolver just before dawn.
The farmer deterred the dangerous wolf with the revolver just before dawn.
The farmer deterred the cunning robber with the raincoat just before dawn.

The man repaired the metal toolbox with the screwdriver after one month.
The man repaired the antique watch with the screwdriver after one month.
The man repaired the antique watch with the malfunction after one month.

The hunter killed the dangerous poacher with the rifle not long after sunset.
The hunter killed the dangerous leopard with the rifle not long after sunset.
The hunter killed the dangerous leopard with the scars not long after sunset.

Peter yelled at the tall protester with the loudspeaker for a long time.
Peter yelled at the slow racehorse with the loudspeaker for a long time.
Peter yelled at the tall protester with the windcheater for a long time.

The scoundrel scratched the yellow car with the aerial after the football match.
The scoundrel scratched the red enamel with the aerial after the football match.
The scoundrel scratched the yellow car with the faults after the football match.

The angry man hit the nasty boy with the crowbar again and again.
The angry man hit the young cat with the crowbar again and again.
The angry man hit the young cat with the stripes again and again.

The surgeon examined the old doctor with the stethoscope after half past five.
The surgeon examined the weak heart with the stethoscope after half past five.
The surgeon examined the old doctor with the temperature after half past five.

The criminal shot the approaching cop with the pistol yesterday late at night.
The criminal shot the approaching dog with the pistol yesterday late at night.
The criminal shot the approaching dog with the collar yesterday late at night.

The doctor wet the brown bottle with the liquid during the examination.
The doctor wet the painful skin with the liquid during the examination.
The doctor wet the painful skin with the rashes during the examination.

The old workman hurt the filthy miner with the pickaxe almost a week ago.

The old workman hurt the nasty poodle with the pickaxe almost a week ago.

The old workman hurt the filthy miner with the stutter almost a week ago.

The annoying boy damaged the iron shelf with the hammer some time ago.

The annoying boy damaged the wine glass with the hammer some time ago.

The annoying boy damaged the wine glass with the cracks some time ago.

The woman dried the iron rack with the towel a little while ago.

The woman dried the wet wound with the towel a little while ago.

The woman dried the wet wound with the blood a little while ago.

The landlady warmed the tiny room with the heater within a couple of minutes.

The landlady warmed the nice soup with the heater within a couple of minutes.

The landlady warmed the nice soup with the garlic within a couple of minutes.

The beggar frightened the strange lunatic with the penknife late at night.

The beggar frightened the small greyhound with the penknife late at night.

The beggar frightened the strange lunatic with the freckles late at night.

The teacher poked the funny girl with the umbrella every time again.

The teacher poked the sandy soil with the umbrella every time again.

The teacher poked the funny girl with the sickness every time again.

The crazy man stabbed the furious knight with the dagger during the unrest.

The crazy man stabbed the fierce bulldog with the dagger during the unrest.

The crazy man stabbed the fierce bulldog with the collar during the unrest.

The teenager washed the soiled sink with the sponge early in the morning.

The teenager washed the injured ear with the sponge early in the morning.

The teenager washed the injured ear with the pimple early in the morning.

The hunter aimed at the old poacher with the firearm during the morning.

The hunter aimed at the grey pigeon with the firearm during the morning.

The hunter aimed at the old poacher with the rabbits during the morning.

The caretaker cleaned the small bucket with the solvent after a while.

The caretaker cleaned the coffee stain with the solvent after a while.

The caretaker cleaned the small bucket with the rubbish after a while.

The lord protected the disputed field with the fence for a long time.

The lord protected the feeding cattle with the fence for a long time.

The lord protected the disputed field with the moles for a long time.

### *Experimental Items from Experiment 2*

The old man poked only the chap with the umbrella a number of times.

The old man poked only the soil with the umbrella a number of times.

The old man poked only the chap with the raincoat a number of times.

The brute frightened only the scoundrel with the penknife late at night.

The brute frightened only the greyhound with the penknife late at night.

The brute frightened only the scoundrel with the freckles late at night.

The farmer injured only the shepherd with the staff a few days ago.

The farmer injured only the sheepdog with the staff a few days ago.

The farmer injured only the sheepdog with the spots a few days ago.

The tourists followed only the naturalist with the jeep during the trip.

The tourists followed only the rhinoceros with the jeep during the trip.

The tourists followed only the rhinoceros with the limp during the trip.

The old man pointed at only the girl with the stick after some time.

The old man pointed at only the star with the stick after some time.

The old man pointed at only the girl with the dress after some time.

The tourist saw only the walker with the telescope early in the afternoon.

The tourist saw only the badger with the telescope early in the afternoon.

The tourist saw only the badger with the infection early in the afternoon.

The tribesman knocked down only the enemy with the cudgel the other day.
The tribesman knocked down only the tiger with the cudgel the other day.
The tribesman knocked down only the tiger with the injury the other day.

The peasant damaged only the shed with the tractor after the hard work.
The peasant damaged only the pane with the tractor after the hard work.
The peasant damaged only the shed with the windows after the hard work.

The explorer found only the hiker with the compass after the hurricane.
The explorer found only the coast with the compass after the hurricane.
The explorer found only the coast with the harbour after the hurricane.

The intruder struck only the hero with the sword during the attack.
The intruder struck only the lock with the sword during the attack.
The intruder struck only the hero with the beard during the attack.

The hunter killed only the poacher with the rifle not long after sunset.
The hunter killed only the leopard with the rifle not long after sunset.
The hunter killed only the leopard with the scars not long after sunset.

Jonathan spotted only the bag with the binoculars after a long search.
Jonathan spotted only the fox with the binoculars after a long search.
Jonathan spotted only the bag with the wristwatch after a long search.

The vandal scratched only the lorry with the aerial after the football match.
The vandal scratched only the paint with the aerial after the football match.
The vandal scratched only the lorry with the faults after the football match.

The African hunter aimed at only the rival with the spear during the pursuit.
The African hunter aimed at only the zebra with the spear during the pursuit.
The African hunter aimed at only the zebra with the wound during the pursuit.

The young ruffian kicked only the girl with the trainers a number of times.
The young ruffian kicked only the ball with the trainers a number of times.
The young ruffian kicked only the girl with the pigtails a number of times.

The adventurer found only the bag with the charts after a long time.
The adventurer found only the sea with the charts after a long time.
The adventurer found only the sea with the whales after a long time.

The man touched only the tramp with the knife after some hesitation.
The man touched only the stain with the knife after some hesitation.
The man touched only the tramp with the fever after some hesitation.

The angry man battered only the boy with the crowbar last week on Sunday.
The angry man battered only the cat with the crowbar last week on Sunday.
The angry man battered only the cat with the stripes last week on Sunday.

The handyman repaired only the toolbox with the pliers right after lunch.
The handyman repaired only the lantern with the pliers right after lunch.
The handyman repaired only the lantern with the damage right after lunch.

The police tracked down only the terrorist with the bulldogs after the chase.
The police tracked down only the narcotics with the bulldogs after the chase.
The police tracked down only the terrorist with the dynamite after the chase.

Peter yelled at only the protester with the megaphone for a long time.
Peter yelled at only the racehorse with the megaphone for a long time.
Peter yelled at only the protester with the moustache for a long time.

A coastguard rescued only the tourist with the motorboat during the patrol.
A coastguard rescued only the dolphin with the motorboat during the patrol.
A coastguard rescued only the dolphin with the poisoning during the patrol.

The trooper destroyed only the tank with the projectile during the fight.
The trooper destroyed only the wall with the projectile during the fight.
The trooper destroyed only the tank with the camouflage during the fight.

The criminal shot only the cop with the pistol yesterday late at night.

The criminal shot only the dog with the pistol yesterday late at night.

The criminal shot only the dog with the collar yesterday late at night.

The caretaker cleaned only the pail with the brush after a while.

The caretaker cleaned only the suit with the brush after a while.

The caretaker cleaned only the pail with the holes after a while.

The man repaired only the toolbox with the screwdriver as quickly as possible.

The man repaired only the printer with the screwdriver as quickly as possible.

The man repaired only the printer with the malfunction as quickly as possible.

The man deterred only the thug with the shotgun just before dawn.

The man deterred only the wolf with the shotgun just before dawn.

The man deterred only the thug with the tattoos just before dawn.

The teenager washed only the basin with the sponge early in the morning.

The teenager washed only the cheek with the sponge early in the morning.

The teenager washed only the cheek with the pimple early in the morning.

The outlaw wounded only the guard with the crossbow during the night.

The outlaw wounded only the horse with the crossbow during the night.

The outlaw wounded only the guard with the whiskers during the night.

The burglar stabbed only the guy with the dagger during the night.

The burglar stabbed only the dog with the dagger during the night.

The burglar stabbed only the dog with the collar during the night.

## REFERENCES


Abney, S. P. (1989). A computational model of human parsing. *Journal of Psycholinguistic Research*, **18,** 129–144.

Adams, B. C., Clifton, C., Jr., & Mitchell, D. C. (1998). Lexical guidance in sentence processing? *Psychonomic Bulletin & Review*, **5,** 265–270.

Altmann, G. T. M., Garnham, A., & Dennis, Y. I. L. (1992). Avoiding the garden path: Eye-movements in context. *Journal of Memory and Language*, **31,** 685–712.

Altmann, G. T. M., Garnham, A., & Henstra, J. A. (1994). Effects of syntax in human sentence parsing: Evidence against a structure-based parsing mechanism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20,** 209–216.

Altmann, G. T. M., & Steedman, M. J. (1988). Interaction with context during human sentence processing. *Cognition*, **30,** 191–238.

Altmann, G. T. M., van Nice, K. Y., Garnham, A., & Henstra, J. A. (1998). Late closure in context. *Journal of Memory and Language*, **38,** 459–484.

Baayen, H. R., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical database (CDROM)*. Philadelphia, PA: Linguistic Data Consortium, Univ. of Pennsylvania.

Binder, K. S., & Rayner, K. (1998). Contextual strength does not modulate the subordinate bias effect: Evidence from eye fixations and self-paced reading. *Psychonomic Bulletin & Review*, **5,** 271–276.

Britt, M. A. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language*, **33,** 251–283.

Britt, M. A., Perfetti, C. A., Garrod, S. C., & Rayner, K. (1992). Parsing in context: Context effects and their limits. *Journal of Memory and Language*, **31,** 293–314.

Brysbaert, M., & Mitchell, D. C. (1996). Modifier attachment in sentence parsing: Evidence from Dutch. *Quarterly Journal of Experimental Psychology*, **49A,** 664–695.

Clifton, C., Jr, Bock, J., & Radó J. (2000). Effects of the focus particle *only* and intrinsic contrast on comprehension of reduced relative clauses. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 591–619). Oxford, UK: Elsevier.

Clifton, C., Jr., & Ferreira, F. (1989). Ambiguity in context. *Language and Cognitive Processes*, **4,** 77–103.

Clifton, C., Jr., Kennison, S. M., & Albrecht, J. E. (1997). Reading the words *her, his, him*: Implications for parsing principles based on frequency and on structure. *Journal of Memory and Language*, **36,** 276–292.

Clifton, C., Jr., Speer, S., & Abney, S. P. (1991). Parsing arguments: Phrase structure and argument structure as determinants of initial parsing decisions. *Journal of Memory and Language*, **30,** 251–271.

Collins, M., & Brooks, J. (1995). Prepositional phrase attachment through a backed-off model. In D. Yarowsky & K. Church (Eds.), *Proceedings of the third workshop on very large corpora* (pp. 27–38). Cambridge, MA: MIT Press.

Crain, S., & Steedman, M. J. (1985). On not being led up the garden-path: The use of context by the psychological processor. In D. Dowty, L. Kartunnen, & A. Zwicky (Eds.), *Natural language parsing* (pp. 320–358). Cambridge, UK: Cambridge Univ. Press.

Crocker, M. W. (1995). *Computational psycholinguistics: An interdisciplinary approach to the study of language*. Dordrecht, The Netherlands: Kluwer.

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition*, **30,** 73– 105.

Cuetos, F., Mitchell, D. C., & Corley, M. M. B. (1996). Parsing in different languages. In M. Carreiras, J. Garcia-Albea, & N. Sabastian-Galles (Eds.), *Language processing in Spanish* (pp. 145–187). Hillsdale, NJ: Erlbaum.

De Vincenzi, M., & Job, R. (1993). Some observations on the universality of the late closure strategy: Evidences from Italian. *Journal of Psycholinguistic Research*, **22,** 189–206.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, **27,** 429–446.

Ferreira, F., & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language*, **25,** 348–368.

Ferreira, F., & Henderson, J. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **16,** 555–568.

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies*. Ph.D. dissertation, University of Connecticut. Indiana University Linguistics Club.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII* (pp. 559–586). Hillsdale, NJ: Erlbaum.

Frazier, L. (1995). Constraint satisfaction as a theory of sentence processing. *Journal of Psycholinguistic Research*, **24,** 437–468.

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 and Language*, **37,** 58–93.

Gilboy, E., Sopena, J.-M., Clifton, C., Jr., & Frazier, L. (1995). Argument structure and association preferences in Spanish and English complex NPs. *Cognition*, **54,** 131–167.

Hemforth, B., Konieczny, L., & Scheepers, C. (2000). Syntactic attachment and anaphor resolution: Two sides of relative clause attachment. In M. W. Crocker, M. J. Pickering, & C. Clifton, Jr. (Eds.), *Architectures and mechanisms for language processing* (pp. 259–281). Cambridge, UK: Cambridge Univ. Press.

Hindle, D., & Rooth, M. (1993). Structural ambiguity in lexical relations. *Computational Linguistics*, **19,** 103–120.

Holmes, V. M., & O'Regan, J. K. (1981). Eye fixation patterns during the reading of relative-clause sentences. *Journal of Verbal Learning and Verbal Behavior*, **20,** 417–430.

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

Konieczny, L., Hemforth, B., Scheepers, C., & Strube, G. (1997). The role of lexical heads in parsing: Evidence from German. *Language and Cognitive Processes*, **12,** 307–348.

Lewis, R. (1999). *Attachment without competition: A computational model of race-based parsing in a limited working memory*. Paper presented at the 12th annual CUNY conference, New York, NY.

Liversedge, S. P., Pickering, M. J., Branigan, H. P., & Van Gompel, R. P. G. (1998). Processing arguments and adjuncts in isolation and context: The case of *by*-phrase ambiguities in passives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **24,** 461–475.

MacDonald, M. C. (1993). The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language*, **32,** 692–715.

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, **9,** 157–201.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, **24,** 56– 98.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, **101,** 676–703.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, **38,** 283–312.

McRoy, S. W., & Hirst, G. (1990). Race-based parsing and syntactic disambiguation. *Cognitive Psychology*, **14,** 313–353.

Merlo, P., Crocker, M. W., & Berthouzoz, C. (1997). Attaching multiple prepositional phrases: Generalized backed-off estimation. In C. Cardie & R. Weischedel (Eds.), *Proceedings of the Second Conference on Empirical Methods in Natural Language Processing*. (p. 149–155). Providence, RI.

Mitchell, D. C. (1987). Lexical guidance in human parsing: Locus and processing characteristics. In Coltheart, M. (Ed.), *Attention and performance XII* (pp. 601–618). Hillsdale, NJ: Erlbaum.

Mitchell, D. C., Corley, M. M. B., & Garnham, A. (1992). Effects of context in human sentence parsing: Evidence against a discourse-based proposal mechanism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **18,** 69–88.

Mitchell, D. C., & Cuetos, F. (1991). The origins of parsing strategies. In C. Smith (Ed.), *Current issues in natural language processing* (pp. 1–12). Austin: Center for Cognitive Science, Univ. of Texas.

Mitchell, D. C., Cuetos, F., Corley, M. M. B., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research*, **24,** 469–488.

Mitchell, D. C., & Green, D. W. (1978). The effect of context and content on immediate processing in reading.

*Quarterly Journal of Experimental Psychology*, **30,** 609–636.

Ni, W., Crain, S., & Shankweiler, D. (1996). Sidestepping garden paths: Assessing the contributions of syntax, semantics and plausibility in resolving ambiguities. *Language and Cognitive Processes*, **11,** 283–334.

Paterson, K. B., Liversedge, S. P., & Underwood, G. (1999). The influence of focus operators on syntactic processing of "short" relative clause sentences. *Quarterly Journal of Experimental Psychology*, **52A,** 717–737.

Pearlmutter, N. J., & MacDonald, M. C. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language*, **34,** 521–542.

Pearlmutter, N. J., & Mendelsohn, A. A. (1999). *Serial vs. parallel sentence comprehension*. Unpublished manuscript.

Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **24,** 940–961.

Pickering, M. J., Traxler, M. J., & Crocker, M. W. (2000). Ambiguity resolution in sentence processing: Evidence against likelihood. *Journal of Memory and Language*, **43,** 447–475.

Pritchett, B. (1992). *Grammatical competence and parsing performance*. Chicago: Univ. of Chicago Press.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, **22,** 358–374.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, **14,** 191–201.

Rayner, K., Garrod, S. C., & Perfetti, C. A. (1992). Discourse influences in parsing are delayed. *Cognition*, **45,** 109–139.

Rayner, K., Kambe, G. A., & Duffy, S. A. (2000). The effect of clause wrap-up on eye movements during reading. *Quarterly Journal of Experimental Psychology*, 53A, 1061–1080.

Rayner, K., & Morris, R. K. (1991). Comprehension processes in reading ambiguous sentences: Reflections from eye movements. In G.B. Simpson (Ed.), *Understanding word and sentence* (pp. 175–198). Amsterdam: Elsevier.

Rayner, K., Pacht, J. M., & Duffy, S. A. (1994). Effects of prior encounter and global discourse bias on the processing of lexically ambiguous words: Evidence from eye fixations. *Journal of Memory and Language*, **33,** 527–544.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice-Hall.

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton, C., Jr. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes*, **4,** SI 21–49.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, **105,** 125–157.

Schütze, C. T., & Gibson, E. (1999). Argumenthood and English prepositional phrase attachment. *Journal of Memory and Language*, **40,** 409–431.

Speer, S. R., & Clifton, C., Jr. (1998). Plausibility and argument structure in sentence comprehension. *Memory & Cognition*, **26,** 965–978.

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **24,** 1521–1543.

Spivey-Knowlton, M., & Sedivy, J. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, **55,** 227–267.

Tabor, W., Juliano, C., & Tanenhaus, M. K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes*, **12,** 211–271.

Tabor, W., & Tanenhaus, M. K. (1999). Dynamical models of sentence processing. *Cognitive Science*, **23,** 491–515.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *The handbook of perception and cognition: Speech, language, and communication* (Vol. 11, pp. 217–262). San Diego CA: Academic Press.

Taraban, R., & McClelland, J. R. (1988). Constituent attachment and thematic role assignment in sentence processing: Influence of content-based expectations. *Journal of Memory and Language*, **27,** 597–632.

Traxler, M. J., Bybee, M. D., & Pickering, M. J. (1997). Influence of connectives on language comprehension: Eye-tracking evidence for incremental interpretation. *Quarterly Journal of Experimental Psychology*, **50A,** 481–497.

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, **35,** 454–475.

Traxler, M. J., Pickering, M. J., & Clifton, C., Jr. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, **39,** 558–592.

Traxler, M. J., Pickering, M. J., Clifton, C., Jr., & Van Gompel, R. P. G. (2000). Architectures and mechanisms for sentence processing: Is syntactic parsing a form of lexical ambiguity resolution? In M. De Vincenzi & V. Lombardo (Eds.), *Cross-linguistic studies of sentence processing* (pp. 149–174). Dordrecht, The Netherlands: Kluwer.

Trueswell, J. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, **35,** 566–585.

Trueswell, J., Tanenhaus, M. K., & Garnsey, S. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language*, **33,** 285–318.

Trueswell, J., 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, and Cognition*, **19,** 528–553.

Tyler, L. K., & Marslen-Wilson, W. D. (1977). The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behavior*, **16,** 683–692.

Van Gompel, R. P. G., & Pickering, M. J. (in press). *Lexical guidance in sentence processing: A note on Adams, Clifton, and Mitchell (1998)*. *Psychonomic Bulletin & Review*.

Van Gompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2000). Unrestricted race: A new model of syntactic ambiguity resolution. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 621–648). Oxford, UK: Elsevier.

(Received December 23, 1998)
(Revision received October 16, 2000)
Published Online June 21, 2001