Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities

Julie A. Van Dyke[a,*] and Richard L. Lewis[b]

[a] *Haskins Laboratories, 270 Crown Street, New Haven, CT 06511-6695, USA*
[b] *University of Michigan, USA*

Abstract

This paper presents the cue-based retrieval theory of parsing and reanalysis and illustrates how this account can accommodate a number of key results about parsing and reanalysis, including effects due to structure, distance, and type of structural change. Three offline experiments and one online experiment permit establishing the locus of these effects as due to properties of the initial parsing processes or to the repair mechanism. Specifically, the data reported here suggest that a structural factor specific to the operation of the parser, *retrieval interference*, affects attachment uniformly across ambiguous and unambiguous sentences and serves to create a limit on successful repair. In addition, these experiments suggest that distance of the head of an ambiguous phrase from its disambiguator affects repair processes—and not attachment processes—independently of the interference effect. These results are interpreted with respect to alternative models of reanalysis, which are contrasted with the cue-based retrieval account, which requires no distinct repair mechanism to account for the current results. A further contribution of this article is to suggest a statistical correction for individual variance in reading rates. Statistical analyses on individual subject data confirmed previous speculations regarding a possible increase in reading rates as subjects move through a sentence. While this individual variation limits fair comparisons of reading times in sentence regions that appear in non-identical serial positions, we demonstrate that such comparisons become meaningful when the appropriate regression analyses have been performed.
© 2003 Elsevier Inc. All rights reserved.

Keywords: Parsing; Reanalysis; Ambiguity resolution; Memory retrieval

Theories of human sentence comprehension in the last 30 years have primarily addressed questions concerning the initial processing of temporary ambiguities, but recently researchers have paid increasing attention to complementary questions about how the human sentence processing mechanism recovers when initial interpretations prove incorrect (Ferreira & Henderson, 1991b; Fodor & Inoue, 1994, 1998; Frazier & Clifton, 1998; Pickering & Traxler, 1998; Schneider & Phillips, 2001; Stevenson, 1998; Sturt, Pickering, & Crocker, 1999; Sturt, Pickering, Scheepers, & Crocker, 2001; van Gompel, Pickering, & Traxler, 2001). Understanding recovery processes is important because the sentence processor's response to pervasive local ambiguities is a function of both the principles that govern the initial processing of the local ambiguity, and the principles that govern the processing of the disambiguating material. For example, a theory of why certain sentences yield

---

* Corresponding author. Fax: 1-203-865-8963.
E-mail address: jvandyke@haskins.yale.edu (J.A. Van Dyke).

difficult *garden path* effects must provide both an explanation of why the incorrect path was initially preferred, and why the recovery was difficult (Frazier & Rayner, 1982).

The current paper addresses these questions in two ways. First, we present a new model of parsing, the cue-based retrieval parser, together with empirical results illustrating the unique predictions this parser makes about the types of syntactic attachments expected to cause difficulty. These data support the chief axiom of cue-based parsing: integrating incoming words into the existing interpretation is limited by *decay* and *retrieval interference*. We will argue that these are pervasive constraints, affecting the processing of ambiguous, unambiguous, and disambiguating input. Secondly, we provide a theoretically neutral, functional analysis of the reanalysis mechanism that provides a framework for evaluating the impact of these constraints on different stages of the reanalysis process. In so doing, we will highlight the extent to which reanalysis mechanisms are distinct from those employed by the sentence processor in handling unambiguous materials (or ambiguous materials disambiguated toward the heavily preferred interpretation). We conclude that much of the work required for reanalysis can be achieved via the cue-based retrieval mechanism, thus eliminating the need for separate principles governing reanalysis.

Cue-based parsing and retrieval interference

Any predictions regarding factors that affect reanalysis require a well-specified model of parsing. Traditionally, these predictions come from the constraints built into the parser, such as limits on constituent length (Frazier & Fodor, 1978; Warner & Glass, 1987), storage constraints (Gibson, 1991, 1998; Kimball, 1973), and rules on structure building or destroying (Frazier, 1987; Gorrell, 1995; Sturt et al., 1999). The parsing model to be described here takes a different approach, in that constraints on parsing success are defined with respect to the ease with which constituents required for the parse, whether during initial interpretation or reanalysis, can be retrieved from working memory (see Van Dyke (2002) and Lewis, Van Dyke, Vasishth, & Nakayama (submitted) for more detailed presentations). In this respect, our model bears similarity to models from the episodic memory literature that focus on the adequacy of the retrieval cue for unambiguously identifying a previously stored target item from among a set of potentially interfering candidates (e.g., SAM, Gillund & Shiffrin, 1984; MINERVA 2, Hintzman, 1988, see Clark & Gronlund, 1996 for a review), as well as models of working memory that focus on retrieval interference and activation decay as the factors limiting performance,

rather than storage capacity (e.g., ACT-R (Anderson & Lebiere, 1998; Anderson & Matessa, 1997)).

For the most part, retrieval models have not been directly applied to problems of language processing. A notable exception is the work of McElree and colleagues (2000; McElree, Foraker, & Dyer (2003)). Using a speed-accuracy tradeoff procedure, they presented data suggesting that working memory structures are retrieved in a manner consistent with cue-based retrieval, and not with a systematic search through verbal working memory. In particular, McElree et al. (2003) showed that the speed of processing structures with increasing distance between two grammatically dependent items, as well as with increasing the structural complexity of the interpolated material, remained constant, suggesting that retrieval speed is unaffected by the position of the retrieved item vis-à-vis its retrieval cues. This would not be the case if the parser executed a step-by-step traversal back through the parse tree to locate items necessary for completing grammatical dependencies, since the length of this search varies with the increased distance and/or complexity of the intervening material. Thus, the McElree et al. results suggest that the completion criteria of the underlying retrieval mechanism are not governed by the grammatical or structural context of particular sentence, but rather is *content-addressable*, meaning that retrieval is carried out via a direct and parallel matching of all possible candidates against the retrieval cues available when the retrieval is initiated (cf. Gillund & Shiffrin, 1984 or Hintzman, 1984 for a mathematic model of this process).

The reliance on retrieval cues inherent in these models suggests an important role for linguistic context in parsing. Indeed, a substantial amount of work already attests to contextual properties affecting sentence processing, such as the structural or semantic biases of critical items which may influence initial attachments (Fodor, 1978; Ford, Bresnan, & Kaplan, 1982; Garnsey, Pearlmutter, Myers, & Lotocky, 1997; MacDonald, Pearlmutter, & Seidenberg, 1994; McElree et al., 2003; Mitchell, Cuetos, Corley, & Brysbaert, 1995; Pickering & Traxler, 1998; Trueswell, Tanenhaus, & Garnsey, 1994; Trueswell, Tanenhaus, & Kello, 1993). From the cue-based retrieval point of view, these factors all comprise different types of cues which may be utilized during the retrieval of required constituents. From the perspective of memory models, these lexical effects are only one possible source of contextual influences on parsing, however. The models suggest it is not simply the identity of the cues themselves that determine retrieval, but the strength of the match between these cues and the other items in working memory. As such, it is necessary to consider how other items in the sentence can contribute to processing breakdown, even when they are not immediately involved in the grammatical dependencies being created. For example, Lewis (1996) presented a

model suggesting that storing more than 2 items under the same syntactic relation can produce *similarity-based interference*, due to a hypothesized working memory constraint on storing more than 2 (or 3) similar items. While this model predicts interference affects due to syntactic properties of the surrounding linguistic material, it differs from standard memory models which focus on the interference generated by a process of matching retrieval cues to previously stored linguistic items (although see (Lewis et al., submitted) for a discussion of how cue-based retrieval models embody the claims of the earlier Lewis (1993, 1996) model.)

More recent empirical support for the role of similarity-based interference has come from Gordon, Hendrick, and Johnson (2001), who provided evidence that relative clauses were more difficult to process when the critical NPs were similar (i.e., both were definite descriptions or else both were proper names) than when they were mixed. In addition, Gordon, Hendrick, and Levine (2002) demonstrated similar effects with cleft structures when subjects were asked to remember a list of either descriptive nouns or names prior to reading the critical sentence. When both the list and the sentence contained proper names (or else descriptive NPs), processing was more difficult than when the list and the sentence were unmatched for type of NP. These results are consistent with the view that comprehension of the clefted sentences requires retrieval of previously stored items in order for them to be integrated within the linguistic representation being constructed. When the necessary items are similar to other items in the memory load, these items will compete with the intended one, creating retrieval interference that will make it more difficult to accurately interpret the sentence.

The evidence in support of a cue-based retrieval mechanism at work in sentence processing has important implications for models of reanalysis, as it appears contradictory to those models which suggest that reanalysis involves a backwards search through previously constructed dependencies in order to diagnose and correct the initial error (e.g., Fodor & Inoue, 1994, 1998; Lombardo (1998)). As such, we endeavor to present an alternative model in which reanalysis depends simply on the processor's ability to retrieve prior constituents or lexical frames necessary for completing the appropriate attachment. Extrapolating from predictions of the memory models, we expect that retrieval, and hence reanalysis, will be fast and efficient when (a) the necessary item maintains a sufficient degree of activation in working memory so as to make it available for retrieval (i.e., it has not decayed beyond a retrieval threshold) and (b) the available retrieval cues can unambiguously identify the necessary item (i.e., there is minimal retrieval interference). Although these claims refer primarily to the operation of the parsing mechanism and its

ability to build structure using a cue-based retrieval mechanism, reanalysis provides an excellent example of how these two factors work together. Consider the ambiguous sentence in (1).

(1) The secretary forgot the student who was waiting for the exam was standing in the hallway.

Prior to the point of ambiguity, the cue-based retrieval parser operates in a serial fashion (i.e., maintaining a single representation (Lewis, 2000)), using knowledge of the grammar to create predictions about expected constituents in the manner of a left-corner parser (cf. Abney & Johnson, 1991). Links between predicted constituents and their lexical realizations are created in response to retrievals initiated and guided by the lexical requirements of the items in the sentence (see Lewis et al. (submitted) for a detailed discussion of cue-based retrieval parsing). We assume that all lexical senses are activated simultaneously (Duffy, Morris, & Rayner, 1988; Gernsbacher, 1990; Kintsch & Mross, 1985; Onifer & Swinney, 1981; Perfetti & Tan, 1998; Rayner & Duffy, 1986; Rayner, Pacht, & Duffy, 1994; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982), and their level of activation varies depending on their frequency of occurrence in the language (Binder & Rayner, 1998; Duffy et al., 1988; Hogaboam & Perfetti, 1975; Neill, Hilliard, & Cooper, 1988). These relative differences determine the initial attachments to be made as well as the probability of retrieving necessary items.

For example, when the word *forgot*, which can take either an NP complement or a sentential complement structure, occurs in sentence (1), the initial attachments will be determined by the relative frequency of the two lexical frames. In this case, this causes an NP complement structure to be created and the phrase *the student who was waiting for the exam* becomes the object of the verb, allowing the alternative lexical frame to decay (Altmann & Gray, 2002; Ferreira & Henderson, 1991b). When the disambiguating word *was* occurs, the parser must break the object link between *forgot* and *student* in order to attach *was* into the existing parse tree. In addition, the NP that was the object of *forgot* must now be reattached as the subject of the IP headed by *was* and the entire new clause must be attached as the complement of the alternative sentential complement-taking sense of *forgot*. Thus, two retrievals are required: one for the alternative sense of *forgot*, the assigner of the sentential complement, and one for the subject itself.

Fig. 1a illustrates the cue-based retrieval of *forgot* initiated by the disambiguating verb *was* (to simplify presentation we show words and the retrieval of words, although it is projected constituents that are retrieved). In this illustration, we have made explicit two of the grammatically relevant features which we assume are associated with each lexical item, the fact that *was* is a verb (V) which can head a sentential complement

+V
+scomp

+V
+vcomp

forgot   student   who   was   exam

was

Retrieval cues
{V,scomp}

(a)

+N
+sing
+Nom

+N
+sing
+Nom
+empty

+N
+sing
+Acc

forgot   student   who   was   exam

was

Retrieval cues
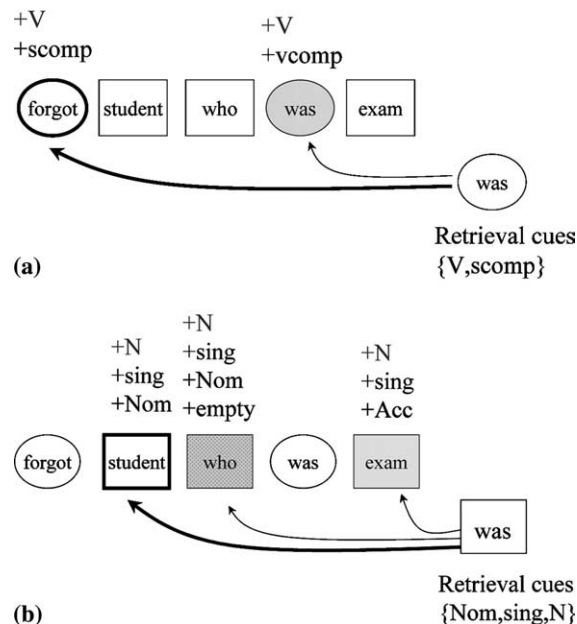{Nom,sing,N}

(b)

Fig. 1. (a) Low interference at verb retrieval for the sentence *The secretary forgot the student who was waiting for the exam was standing in the hallway.* Partial matches are indicated by items having the same shape as the target, shown with a darkened border. The amount of interference provided by intermediate items is represented by the darkness of the shading, with darker shading suggesting more interference. (b) Low interference at subject retrieval for the sentence *The secretary forgot the student who was waiting for the exam was standing in the hallway.*

(scomp).[1] These are the most relevant features for determining where *was* will be attached into the existing parse tree, because they specify the features of the attachment site, and hence they comprise the most distinguishing cues for retrieving the appropriate attachment host, which in this case are matched only by the alternative sense of the verb *forgot*. Notably, these features are partially overlapping with those for the intermediate verb *was*, which means that some interfer-

ence is contributed by this item. However, since the overlap is only partial, the interference is minimized because it can be easily distinguished by the mismatched features. This is an illustration of an important property of the cue-based, parallel retrieval: intermediate attachment sites are not considered if an incomplete match with retrieval cues can rule them out. Thus, unlike parsing models which require a step-by-step traversal back through the parse tree in order to find the appropriate attachment site (e.g., Fodor & Inoue, 1994, 1998), this model uses the cues of the disambiguating item to directly select the attachment site from among the possibilities in the parse tree. In this case, the cues provided by the disambiguating item *was* are sufficient to identify the source of the ambiguity; however, retrieval can only succeed if the alternative sense of *forgot* has not decayed to a level that makes it unavailable for retrieval. We assume for the purposes of this example that this is not the case, but we will see in the experiments below that in fact decay does limit reanalysis in sentences such as those in (1).

Once the attachment site is identified by retrieving the appropriate sense of *forgot*, the disambiguating *was* can be attached as the head of its sentential complement. We assume that the reactivation of the sentential complement subcategorization frame of *forgot* triggers deactivation of the competing NP-complement frame through lateral inhibition (a process at work in the initial lexical access as well, and so not a process specific to reanalysis). The deactivation of the NP-complement frame—and its associated complement relation to *student*–thus makes *student* available to serve as the subject of *was*.[2] This is the other retrieval that must occur, illustrated in Fig. 1b, which also occurs with little distraction from other items in the tree because there is only one NP which matches all the retrieval cues provided by *was*. Notably, there is no need to reparse the detached NP headed by *student*, but rather the entire subtree for *the student who was waiting for the exam* is now available for reattachment. This allows the cue-based retrieval parser to realize the same efficiencies

---

[1] Certainly there are many more features which these items share, however since they are shared they do not help distinguish between possible attachment sites and therefore we leave them inexplicit in our examples. Claims about the operation of the cue-based retrieval parser do not depend on a specific set of features, or even a specific set of distinguishing features, as it is only the match between features of the retrieval cue and features of the stored items which govern its operation. While the descriptions of our features have been influenced by the generative grammar tradition, we expect our predictions to be robust against other features sets suggested by other grammars (i.e., HPSG, Minimalism, LFG, etc.) What is critical is simply that the feature set capture the distinctions necessary for identifying grammatical dependencies within the sentence.

---

[2] Christianson, Hollingworth, Halliwell, and Ferreira (2001) have presented evidence that initial incorrect analyses are not entirely abandoned, resulting in a situation where the NP *the deer* in a sentence such as *While the man hunted the deer ran into the woods* may be simultaneously understood as the object of *hunted* and the subject of *ran*. Such a result is compatible with the model presented here if we understand the deactivation of the initial incorrect syntactic relation to be less than complete, perhaps being determined by the strength of the match between the retrieval cues of the verb and the misanalyzed NP. This would account for the results of Christianson et al. which suggest that the plausibility of the final interpretation and the length of the NP (i.e., the duration of the processor's commitment to the incorrect interpretation) affect the probability of preserving the incorrect interpretation.

as dynamic programming parsers that store and re-use parses of well-formed substrings (e.g., Earley, 1970). Furthermore, it does so without explicitly marking points of ambiguity because the cues from the disambiguating item can identify these points during retrieval.

It is possible, however, that the cues provided by a disambiguating item will not be sufficient to select the appropriate attachment site because of interference from intermediate constituents. For example, consider sentence (2), which contains a highly interfering ambiguous region:

(2) The secretary forgot the student who knew the exam was important was standing in the hallway.

As in the previous example, in order to attach the disambiguating word *was* into the tree, the parser must retrieve both the NP *student* and the VP *forgot*. Figs. 2a and b illustrate how the interference provided by the intervening material makes these retrievals difficult. As illustrated in Fig. 2a, the VP *knew* provides substantial interference for retrieving *forgot* (in fact, it is a perfect match to the retrieval cues; only serial position distinguishes the verbs), while the verb *was* contributes interference by virtue of its category type. For the subject retrieval, Fig. 2b illustrates that the features {N, singular,

subject} are shared by the desired subject *student*, but also the relative clause pronoun *who* (as well as perhaps the subject trace, not shown here) and the intervening subject *exam*. Thus, although the disambiguating verb *was* unambiguously specifies a singular subject, these cues do not unambiguously identify the correct subject, and this is part of what makes the attachment—and hence the reanalysis—difficult. To see how interference in this sentence compares with that in sentence (1) above, consider the number of shared features with the target items *forgot* and *student* as a rough metric for the amount of interference provided by the ambiguous region in this sentence. For the sentence in (2) there are nine shared features, $\{N_{who}, N_{exam}, sing_{who}, sing_{exam}, subj_{who}, subj_{exam}, V_{knew}, V_{was}, scomp_{knew}\}$, while in sentence (1) there are only six, $\{N_{who}, N_{exam}, sing_{who}, sing_{exam}, subj_{who}, V_{was}\}$, despite the fact that the two sentences contain the same number of words and the same number of nouns in the ambiguous region. Based on this, the retrieval interference theory predicts that in a sentence such as (1), it will be less difficult to repair the initial misattachment of *student* than in a more highly interfering structure such as that in (2).

A functional decomposition of reanalysis costs

The prediction of the cue-based retrieval parser about the role of retrieval interference in reanalysis stems from claims about how initial attachments are made (i.e., via retrieval), and not about properties of a specialized reanalysis mechanism. To understand more clearly how the attachment mechanism plays a role in reanalysis, we find it useful to decompose reanalysis into sub-functions so that we can explore how specific variables (e.g., retrieval interference, decay, etc.) affect different components of the reanalysis process. Before proceeding with this task decomposition, however, it is necessary to recognize that any decomposition we propose will be subject to debate. There are two reasons for this. First, as with any process decomposition, there are many levels of abstraction available depending on the granularity of the operations we assume as primitives. Second, different assumptions about the nature of the "first-pass" parser necessarily change the way reanalysis must proceed. Thus, in a serial parser where only one interpretation is pursued at a time, reanalysis requires discarding or repairing incorrect structure and building new, correct structure (Frazier, 1978, 1987; Frazier & Fodor, 1978; Frazier & Rayner, 1982; Lewis, 2000). In contrast, parallel processing models (Garnsey et al., 1997; Gibson & Pearlmutter, 2000; MacDonald et al., 1994; Trueswell et al., 1993) suggest that competing interpretations at all points of ambiguity are maintained simultaneously. In these models, the costs associated with reanalysis are the costs incurred by switching from
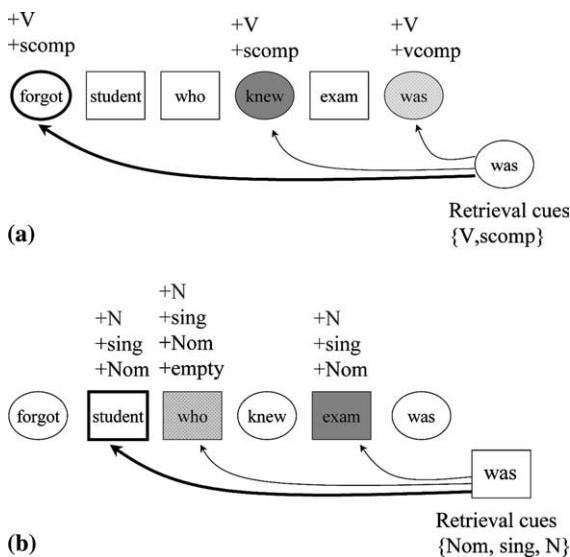

Fig. 2. (a) High interference at verb retrieval for the sentence *The secretary forgot the student who knew the exam was important was standing in the hallway.* Partial matches are indicated by items having the same shape as the target, shown with a darkened border. The amount of interference provided by intermediate items is represented by the darkness of the shading, with darker shading suggesting more interference. (b) High interference at subject retrieval for the sentence *The secretary forgot the student who knew the exam was important was standing in the hallway.*

the preferred interpretation to the non-preferred, but already available interpretation. No structure building during reanalysis is required in these models, merely structure deactivation and reactivation. Thus, the account of reanalysis is different depending on which parsing model the human sentence processing mechanism is assumed to resemble more closely.

Our functional decomposition is thus driven partially by a desire to decompose the task in a way that makes contact with as much previous research as possible. We have identified four functions that we believe are minimally necessary to effect reanalysis. While specific models of parsing and reanalysis may implement these functions differently, all of the functions will be present in some way because of the inherent requirements of the task.

We will motivate the functions via an example in the context of serial parsing and syntactic structure building, and then generalize it to encompass parallel parsing and the building of semantic representations. Consider the example in (3).

(3) When the boys strike the dog kicks.

The initial, but eventually incorrect parse is given in Fig. 3a, while the final correct structure (i.e., the goal of
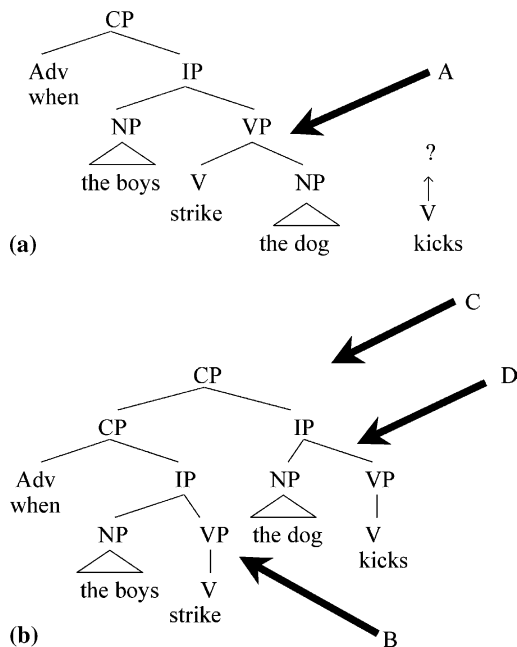
reanalysis) is given in Fig. 3b. In order to make the transition from Fig. 3a to Fig. 3b under a serial parsing view, the HSPM must:

- Identify the locus of the nonfelicitous link (function A);
- Detach that link (function B);
- Switch lexical frames from the transitive sense of *strike* to the intransitive sense (function C);
- Make the correct attachment (function D).[3]

According to the parallel view, both alternative constructions exist, but one is more preferred (e.g., more activated) than the other. Moreover, this will be true at each and every point of ambiguity, so the HSPM is faced with the same problem as seen in the serial model: identifying exactly which ambiguity is the locus where the shift must occur (A). Once the error has been found, the incorrect commitment must be retracted. In a serial model, this means detaching the link, but in a parallel model this calls for adjusting the preferences so as to make the initially preferred structure non-preferred (B). Function (C) in the serial model consists of the reactivation of non-preferred lexical frames; in the parallel model this occurs by adjusting the preferences so as to make the correct structure the one most activated. Function (D) calls on both models to evoke the standard machinery for attaching new items into the tree—in this case the disambiguating constituent itself. Thus, although the content of these operations is a bit different because the parallel model already has the correct alternative parse available, at some level of abstraction both models require the same functional processes in order to successfully complete reanalysis.

Reanalysis cost has traditionally been thought of in terms of variables affecting functions A, B, and C (which together can be thought of as *reanalysis proper*, to distinguish them from the attachment operations (function D) that must happen regardless of whether there was a misanalysis). Until fairly recently, most research has not attempted to localize effects to specific components. A notable exception is the work of Fodor and Inoue (1994, 1998) who have suggested that the major factor determining ease of recovery is the difficulty of diagnosing the grammatical error in the initial structure (function A). In their model, diagnosis happens as a step-by-step backtracking through grammatical dependencies in the incorrect partial parse tree, which if successful leads to a cost-free "cure" (functions B and C) of the incorrect parse. The theory explicitly claims that reanalysis difficulty arises in function A—not B or C. The evidence used to support this model come from manipulations of

Fig. 3. (a) Initial analysis of the sentence *When the boys strike the dog kicks.* Function A indicates that a diagnosis must be done, identifying the incorrect link as the attachment of *the dog* as object of *strikes*. (b) Final analysis of *When the boys strike the dog kicks.* Function B highlights the incorrect link that must be retracted. Function C indicates the structure that must be created upon retrieval of the correct lexical frame of *strike*. Function D suggests the attachment of the disambiguating word into the now reanalyzed parse tree.

---

[3] Note that these functions need not occur in the order in which they are presented here (cf. Fodor & Inoue (1998) and Lewis (1998)) for serial reanalysis models in which attachment occurs first.

the *disambiguating region* (the diagnosis "cue"), which can make a given misanalyzed ambiguity either easy or difficult to diagnose and thus repair (see also Meng & Bader (2000)). There are both important commonalities and differences between the diagnosis model and the cue-based retrieval model, to which we will return in the discussion.

Most studies of reanalysis have investigated properties of the ambiguous region that may make reanalysis more difficult (without necessarily specifying which of the components A–D are affected). For example, Frazier and Rayner (1982) manipulated the length of the ambiguous region in early and late closure sentences and found increased reading times in the disambiguating region for long regions as compared to shorter ones. Subsequent studies have attempted to determine which aspects of the length of the ambiguous region are responsible for the slow down.

Warner and Glass (1987) suggested that it was the number of constituents in the ambiguous region that caused the difficulty. Using sentences like those in (4a and 4b), they argued that as long as the ambiguity is resolved within three constituents, as in sentence (4a) where the disambiguation comes immediately after the ambiguous word *the dog*, then recovery will be successful. If, however, there are more than three constituents intervening between the ambiguity and the disambiguating material, recovery fails and the sentence is perceived to be ungrammatical. This occurs in (4b) where *man* continues to be ambiguous until the fourth constituent, *strikes*, occurs to disambiguate.

(4a) When the boys strike the dog kicks.
(4b) Before the boy kicks the man the dog bites strikes.

One problem with this analysis is that sentence (4b) not only has a longer ambiguous region than (4a), but it also contains a second ambiguity in the reduced relative clause *the man the dog bites*. To clarify these results, Ferreira and Henderson (1991b) (Experiment 1) compared processing in sentences like (4b) with those as in (4c) and found that in addition to length, that additional ambiguities in the ambiguous region make reanalysis difficult. This conclusion is further supported by Tabor and Hutchins (2003) who suggest that ease of reanalysis is affected by the number of incorrect commitments made during initial parsing, which in terms of Component A, means that diagnosing the initial error is even more difficult because there are more opportunities for incorrect attachments in the initial parse.

(4c) Before the boy kicks the man that the dog bites strikes.

A second concern about the proposal that length affects the ease of reanalysis is that it is unclear whether this effect is due to the number of intervening constitu-

ents or the accompanying increase in syntactic complexity, which is correlated with length in the Warner and Glass materials. This issue was addressed by Ferreira and Henderson (1991b) (Experiment 3), who attempted to distinguish length from structure effects by controlling for the number of words in the ambiguous region and varying the syntactic complexity as in (5a and 5b) below.

(5a) When the men hunt the birds that cheetahs eat typically scatter.
(5b) When the men hunt the birds with bright plumage typically scatter.

Sentence (5a) has an ambiguous structure similar in complexity to Warner and Glass' long early closure sentences, while (5b) is argued to be less complex because it contains fewer non-terminal syntactic nodes. Ferreira and Henderson suggested that if it was length per se that affected recovery, then there should be no difference between these sentences, but if the syntactic complexity mattered, then (5a) should be more difficult. They found no difference between the two sentences, concluding that intervening syntactic structure did not affect reanalysis. The experiments to be reported here return to this question, and provide data suggesting that structure does, in fact, play a role.

In a further investigation into which aspects of the ambiguous region are problematic, Ferreira and Henderson (1991a, 1991b) (Experiment 4) compared sentences where length increased the distance between the head of the ambiguous phrase and the disambiguating region and those where it did not. Controlling for overall length of the ambiguous region as well, they compared sentences like (6a), where the head is adjacent to the disambiguating material, to those in (6b).

(6a) While the boy scratched the big and hairy dog yawned loudly.
(6b) While the boy scratched the dog that is big and hairy yawned loudly.

They found that subjects had significantly higher error rates in judging sentences like (6b) to be grammatical as compared with sentences like (6a). They suggested that a longer distance between the head of the ambiguous region (i.e., *dog*) and the disambiguating word (i.e., *yawned*) affects reanalysis because the incorrect thematic role (i.e., *dog* as object of *yawned*) is assigned earlier and the correct thematic argument structure of yawned as an intransitive verb decays as the sentence continues. When the disambiguating information informs the parser that the initial assignment is incorrect, the proper assignment is more difficult to make because the intransitive thematic structure is difficult to retrieve. In contrast, when the head is adjacent to the disambiguating word, there is no opportunity for

the correct but non-preferred thematic structure to decay and the reanalysis can proceed without difficulty. Thus, while this account implicates length in diagnosing the source of the incorrect attachment, it also suggests a role for Component C in that lexical frames which are necessary for the correct analysis may be more difficult to access if the length of the ambiguous region has allowed them to decay enough to make them difficult to retrieve (see also Ferreira & Henderson (1991a)).

Sturt et al. (1999) have suggested that the type of structural change required during reanalysis presents an additional source of difficulty, and perhaps an alternative account for the Ferreira and Henderson effects. Specifically, any restructuring which requires the retraction of existing commitments (function B) is predicted to be costly. This claim is similar to that proposed by Fodor and Frazier (1980) in the principle of Revision-as-a-Last-Resort (RALR), which asserts that the parse built so far is not to be changed unless there is no other way of proceeding (Fodor & Frazier, 1980; Fodor & Inoue, 2000). A similar claim has been suggested by Frazier and Clifton (1998), who argue that the human sentence processor's revisions are governed by a principle of Minimal Revisions, which states that no unnecessary revisions are made and that when revisions are necessary, they should be made in a way that maintains as much as possible of the already assigned structure.

Sturt et al. (1999) note that all of the reanalyses necessary in the sentences tested by Ferreira and Henderson (1991a, 1991b) require abandoning the structural relationship of the matrix verb as dominating the ambiguous NP (as in Fig. 3a) in favor of a structure in which this NP is positioned higher in the tree (Fig. 3b). In a D-theory or minimal commitment framework (Gorrell, 1995; Marcus, Hindle, & Fleck, 1983; Weinberg, 1993), this type of raising is more costly than one in which the matrix verb can take either an NP complement or a sentential complement, as in sentence (7), illustrated in Fig. 4. This is because the initial assignment of *the dog* as being dominated by the verb *know* does not have to be abandoned entirely, but merely revised in a way that preserves this dominance relation.

(7) The boys know the dog was kicking.

To evaluate this prediction, Sturt et al. (1999) compared ambiguous and unambiguous versions of sentences containing verbs which (a) could either take an NP complement or an S complement, and thus required a reanalysis such as that in Fig. 4; and (b) could either take an NP complement or no (Zero) complement, requiring a reanalysis as in Fig. 3b. Calling these verbs NP/S and NP/Z, respectively, they present data that shows that reanalysis with NP/S verbs is less difficult for the parser than reanalysis with NP/Z verbs. In particular, the garden-path effect, calculated by subtracting the reading times in the critical region in unambiguous sentences from reading times for ambiguous versions of the same sentences, for NP/S sentences like (8) was more than four times (400 vs. 87 ms) that of the garden-path effect for NP/Z sentences like in (9).

(8a) NP/S ambiguous:
The Australian woman saw the famous doctor had been drinking quite a lot.
(8b) NP/S unambiguous:
The Australian woman saw that the famous doctor had been drinking quite a lot.
(9a) NP/Z ambiguous:
Before the woman visited the famous doctor had been drinking quite a lot.
(9b) NP/Z unambiguous:
Before the woman visited, the famous doctor had been drinking quite a lot.

To further examine the trade off between construction type and Ferreira and Henderson's claims about the role of head position as a source of reanalysis cost, Sturt et al. also compared NP/S and NP/Z sentences where the head of the ambiguous NP was either immediately adjacent to the disambiguating information or separated from it by a relative clause, as in (10) and (11).

(10a) NP/S adjacent:
The Australian woman saw the famous doctor had been drinking quite a lot.
(10b) NP/S nonadjacent:
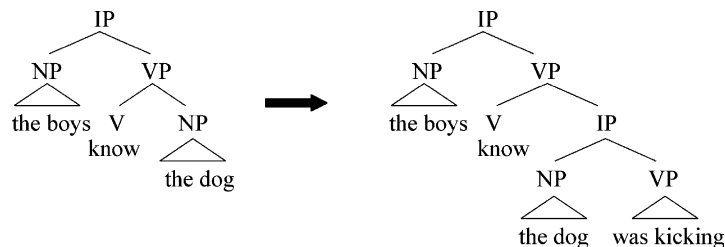The Australian woman saw the doctor who was famous had been drinking quite a lot.

Fig. 4. Reanalysis of NP/S sentence *The boys know the dog was kicking* in which *the dog* is pushed lower in the tree, but still maintains its structural relation as being dominated by the verb *know*.

(11a) NP/Z adjacent:
Before the woman visited the famous doctor had been drinking quite a lot.
(11b) NP/Z nonadjacent:
Before the woman visited the doctor who was famous had been drinking quite a lot.

While they replicated their previous effect of construction type, they found no effect of head position in these sentences. Although Sturt et al. collected reading times, which cannot be directly compared with the Ferreira and Henderson results, their failure to produce the head position effect led them to suggest that the original result was due to an artifact of the RSVP method, which encouraged subjects to rely on superficial clues, such as whether the main verb was close to a word that could function as the head of its subject, when making their judgments. This explanation seems unlikely, however, because in a follow up experiment using eye-tracking methodology with whole-sentence presentation, Ferreira and Henderson (1993) were able to replicate their finding of a head-position effect with grammaticality judgments (see also Bailey & Ferreira (in press) who found head position effects for both NP/Z and NP/S constructions with manipulations of both number of words and presence of disfluencies in the ambiguous region). The results from Ferreira and Henderson (1993) study were not entirely conclusive, since they were unable to find any clear evidence for a head position effect in their eye-movement data. They suggested that this null result is due to a gradual decrease in reading rate as subjects move through the sentences, which may also account for the Sturt et al.'s null result. Since the disambiguating region in sentences such as (12a) occurs comparatively earlier than in sentences such as (12b), the per character reading rate at the disambiguating word in the short conditions will be longer than in the conditions with a long ambiguous region.

(12a) Because Jeff walked THE DOG got upset.
(12b) Because Jeff walked THE BIG AND HAIRY DOG got upset.

This will result in longer total reading times and fixation durations for an early head, as in (12a), whereas the same measures will be shorter for later heads. Thus, any effect of head position in reading times is masked by this decreasing reading rate. Some evidence is available to support this conclusion (cf. Trueswell et al. (1993) and Ferreira & Clifton (1986) for similar results), although they did not perform statistical analyses of this effect. Vasishth (2002) did perform analyses for this effect in self-paced reading times for Hindi and found no effect. However, these analyses were done over all subjects, so individual variation in reading rates may have masked the effect. Van Dyke (2002) did analysis on individual subjects' data and found that 38 out of 48 subjects in a self-paced reading study did show the gradual decrease as they moved through the sentence. Experiment 4 reported here presents analyses using the same method employed by Van Dyke (2002), and together with results from our Experiments 1–3 suggest that the original head position effect is real and does affect reanalysis.

A key contribution of the current research is to investigate costs on reanalysis due to function D, attachment operations, which are implicated in repairing the incorrect parse. There have been few studies to explore these effects, and those which have been done have examined non-structural effects such as verb bias and plausibility effects. However, it is often difficult to ascertain whether the demonstrated effects necessarily reflect properties of the parsing mechanism (function D) or those of the reanalysis mechanism (functions A–C), or possibly both. For example, Garnsey et al. (1997) showed that verb bias affects processing in the disambiguating region in NP/S ambiguities, as in (13a and 13b).

(13a) The photographer accepted the money could not be sent yet.
(13b) The job applicant believed the interviewer had been dishonest.

Processing in the disambiguating region (which forced the sentential complement form) was slower in sentences like (13a), which used verbs that are biased to the NP-complement structure, compared to sentences like (13b), which used verbs biased to the sentential complement structure. This is consistent with an account in which verb bias affects the initial interpretation of the ambiguous constituent, although it is also possible that the verb bias effect stemmed from a reluctance to retract commitments (function B) which are highly supported by preferences of the verb (cf. Stevenson (1998) for a computational model with this property). To take another example, Pickering and Traxler (1998) found that the plausibility of both the initially preferred and ultimately required analyses affected processing difficulty. In particular, increasing the plausibility of the initially preferred interpretation increased difficulty in the disambiguating region, while increasing the implausibility of the initial interpretation increased difficulty at the ambiguity, but *decreased* difficulty in the disambiguating region.

While these results are inconclusive about the locus of effects associated with verb bias and plausibility, they nevertheless suggest that the attachment mechanism may be sensitive to specific lexical and semantic properties of the ambiguity. In the experiments to be presented here, we are concerned with structural properties of the ambiguous region and how they affect attachment as well as reanalysis. In particular, we have seen that the cue-based parser is hindered in contexts where there is a high amount of retrieval interference generated by

structurally similar items in the existing parse ((Lewis, 1996), see also (Gordon et al., 2001)). This is in contrast to the results of Ferreira and Henderson (1991a, 1991b) discussed previously, in which structural properties of the ambiguous region appear not to play a role in reanalysis. Thus, as an initial step in specifying the role of syntactic structure in reanalysis, Experiment 1 simultaneously replicates and challenges these data. Experiments 2 and 3 turn to the task of establishing the locus of structure and head position effects on repair and attachment processes. In order to be clear about where these factors have their effects, we will refer to the operations that implement function D as *attachment* operations, and the operations that implement functions A–C as *repair* operations for the remainder of the paper. Experiment 4 will confirm the distinction observed in Experiments 2 and 3 with online data, as well as providing statistical evidence addressing the debate over head position results observed in Ferreira and Henderson's initial study. Finally, we will show that a retrieval interference theory of working memory provides a unified account of the constraints on reanalysis uncovered in the present experiments and much other prior work.

Experiment one

As described previously, Ferreira and Henderson (1991b) failed to find an effect of structure using sentences like those in (5a and 5b), repeated here as (14a and 14b).

(14a) When the men hunt the birds that cheetahs eat typically scatter.
(14b) When the men hunt the birds with bright plumage typically scatter.

We suggest that the reason for this is that the structures they used did not have the properties to which reanalysis is sensitive—that is, their structures were not sufficiently *interfering* and therefore they presented no difficulties for the cue-based parser. This can be seen in Fig. 5a, which illustrates the retrieval for the verb *hunt*. Here we find only a single interfering feature, $\{V_{eat}\}$, because although the intermediate verb shares its category with *hunt*, it can not also take a sentential complement (scomp). Fig. 5b illustrates the interference in the ambiguous region for the retrieval of the subject *birds*, for which five interfering features can be counted as follows: $\{N_{cheetahs}, sing_{cheetahs}, subj_{cheetahs}, N_{trace}, N_{sing}\}$. This totals to 6, which is the same number of interfering features counted in the low interference sentences such as (1), repeated here as (15a) and therefore we predict both 14a and 15a to cause little difficulty for the cue-based retrieval parser. In contrast, based on the prediction that retrieval interference affects reanalysis because items required for the repair
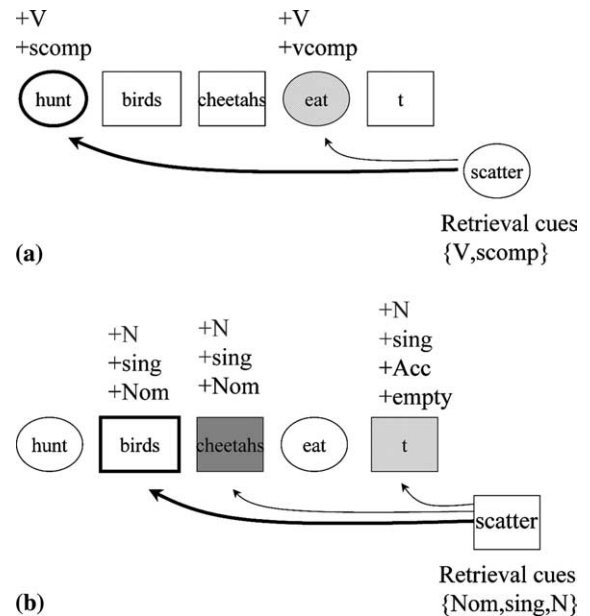

Fig. 5. (a) Interference at subject retrieval for Ferreira and Henderson's complex sentence *When the men hunt the birds that cheetahs eat typically scatter*. (b) Interference at verb retrieval for Ferreira and Henderson's complex sentence *When the men hunt the birds that cheetahs eat typically scatter*.

cannot be retrieved, we predict highly interfering sentences such as (2), repeated here as (15b) to be more difficult to reanalyze.

(15a) The secretary forgot the student who was waiting for the exam was standing in the hallway.
(15b) The secretary forgot the student who knew the exam was important was standing in the hallway.

To evaluate this hypothesis, Experiment 1 directly compared high interference and low interference sentences with those Ferreira and Henderson (1991b) used in their test of syntactic complexity. We expected to replicate Ferreira and Henderson's failure to find an effect of complexity in their sentences, but to see that high interference sentences were more difficult to process than low interference sentences.

Method

Subjects

Twenty-four subjects from the University of Pittsburgh participated in the experiment for partial course credit. All subjects were native speakers of American English.

Materials

Since we were specifically interested in syntactic reanalysis in these experiments, we used only structures

violating the property of late closure, making reanalysis necessary to obtain the globally correct interpretation (Frazier, 1987). The exact stimuli used in Ferreira and Henderson (1991b) (F&H) were obtained, and the early closure versions of these constituted one half of the stimuli used here. The other half contained structures such as those illustrated in sentences (15a) and (15b) above, which we will refer to as our (Van Dyke & Lewis, or V&L) sentences. For purposes of explication, we classify each sentence type in three categories, labeled as short, long/complex, and long/non-complex. Table 1 gives examples from all six experimental conditions.

The short ambiguous region condition for both the F&H sentences and V&L sentences contained only the noun phrase following the first verb of the sentence. The ambiguous region in the long/complex condition was constructed according to each of the two predictions regarding what types of syntactic complexity should increase reanalysis difficulty. In the case of Ferreira and Henderson, this meant object relative clauses, which increased the number of syntactic nodes in the ambiguous region; in the case of V&L sentences, this meant center-embedded subject-relative clauses as described above. The ambiguous region in the long/non-complex sentences was constructed to test the two theories' predictions. In the F&H sentences, this meant prepositional phrases in the ambiguous region (i.e., fewer syntactic nodes with equal number of words); in the V&L sentences, this meant non-embedded subject-relative clauses. In both cases, the total number of words and the number of noun phrases in the ambiguous region was controlled, although by necessity the ambiguous regions in V&L sentences were longer than those in the F&H sentences. The long versions of both F&H and the V&L sentences contained only two noun phrases in the ambiguous region. In addition, all main verbs used in V&L

sentences were checked for their complement bias using the University of Pennsylvania TREEBANK-3 statistics (Marcus, Santorini, & Marcinkiewicz, 1993) on the Brown corpus (Francis & Kucera, 1982). We wanted to ensure that the verbs used in our materials were more biased towards the NP complement reading as compared to the sentential complement reading so that misanalysis was likely to occur. Table 2 gives the verb biases for each of the verbs used in the interfering materials, where NP (and Scomp) bias was calculated as a ratio the number of instances the verb occurs with an NP complement (and sentential complement) over the total number of instances it occurs as a verb in the corpus. Each verb was used twice in our materials set.

Table 2
Verb bias statistics for interfering type experimental sentences

Word	S-comp bias	NP-obj bias
recognized	.12	.86
found	.24	.71
proclaimed	.11	.67
read	.04	.88
wrote	.12	.80
denied	.17	.83
heard	.16	.81
felt	.30	.54
saw	.08	.88
proved	.12	.74
forgot	.21	.65
knows	.50	.38
remember	.28	.67
understand	.30	.62
notice	.35	.63
reported	.25	.74
recalled	.50	.50
envisioned	.00	1.00

Table 1
Sample items, Experiment 1

Ferreira and Henderson (1991a, 1991b) (F&H) sentences
Short region
When the men hunt the birds typically scatter

Long, complex region
When the men hunt the birds that cheetahs eat typically scatter

Long, non-complex region
When the men hunt the birds with bright plumage typically scatter

Van Dyke & Lewis (V&L) sentences
Short region
The frightened boy understood the man was paranoid about dying

Long, complex region (high interference)
The frightened boy understood the man who said the townspeople are dangerous was paranoid about dying

Long, non-complex region (low interference)
The frightened boy understood the man who was swimming near the dock was paranoid about dying

Following the procedures used by Ferreira and Henderson (1991b), thirty-six experimental items in the V&L structures were constructed so that each item had versions in each of the three conditions (see Table 1). Each subject saw only one version of any item, but saw all six conditions of the experiment (i.e., three conditions each for the F&H sentences and the V&L sentences). Across subjects, each version of each sentence was seen an equal number of times. Items were presented in a blocked random order such that every experimental item was separated by two filler items of different syntactic constructions.

There were 72 filler items, designed to be appropriate matches for several aspects of the experimental sentences and to encourage minimal attachment so that subjects would indeed garden path on the experimental sentences. Half of the filler items used verbs with high NP-complement bias and no occurrence with sentential complements, according to the Penn Treebank-III (Marcus et al., 1993). In order to match the structures of our interfering items, we used subject relative clauses as objects of these verbs, (e.g., *The informed citizen elected the candidate who spoke in Arkansas and Pennsylvania.*) Thus, the direct object of these sentences was similar to the objects in our long/low interference items except that there was no ambiguity of attachment and no reanalysis required. Of the other half, two-thirds were transitive sentences with adjective- and/or preposition-modified subjects and preposition modified objects (e.g., *The large hospital with budget problems fired the doctor with the least experience.*) The last sixth of the fillers were multiple clause transitive sentences designed to be long (e.g., *The ski-instructor warned the students of the icy conditions but that didn't prevent them from taking to the slope anyway.*) These were included to discourage subjects from using length as part of their judgment criteria for ungrammatical sentences.

After constructing this set of 72 filler items, we made half of them ungrammatical via the addition of one or two words at random points in the sentences. These were included so as to discourage subjects from adopting any superficial strategies for judging certain structures (especially our complex structures) as ungrammatical.

Procedure

Before the experiment began the subject and the experimenter discussed a list of 10 example sentences, four of which were grammatical. The experimenter explained that for a sentence to be ungrammatical, it should either be missing words or have too many words. This was illustrated with a sentence like *The police gave the citizen who he caught driving too fast on the parkway* or *The student was practicing reviewed his homework.* It was also explained that none of the sentences in the experiment would have punctuation and that subjects should not use this as a reason for judging a sentence ungrammatical.

The experiment was delivered using the MEL2 Professional experimental package (Schneider, 1995) and was run on Pentium-class personal computers. Sentences were presented in a non-cumulative, self-paced, moving window format, where each sentence was presented one word at a time. After reading a review of the instructions they had just discussed with the experimenter, subjects pressed the spacebar to begin the experiment. A fixation cross appeared at the left-most position of the center line of the screen, indicating where the first word would appear. As the subjects continued to press a button with their right hand, each new word would appear and the previous word was replaced by a series of underscores in place of each letter. Thus, the subjects were unable to see the words they had already read in the sentence. At the end of the sentence, the screen was cleared and a question "Is this sentence grammatical?" appeared. Subjects were instructed to indicate their answer with "Yes" or "No" keys designated as the "1" and "2" keys on the keyboard. When they had made their answer, they were instructed to press the space bar when they were ready to begin the next sentence. Before the actual experimental sentences, subjects went through a series of six practice sentences delivered in the same fashion as described here in order to familiarize themselves with the keyboard and the presentation sequence.

Design and analysis

The experiment was analyzed as a 2×3 factorial repeated-measures ANOVA. The first variable was sentence type (i.e., F&H vs. V&L). The second variable was the three different ambiguous region types described above, (i.e., short, long/complex, and long/non-complex). Separate ANOVAs were done for each sentence-type to test for effects between the ambiguous region conditions.

Results

Table 3 shows the percentage correct for the three conditions for each sentence type.

Overall, we found effects of both sentence type and region type. Our sentences were judged grammatical more often than F&H sentences (59% vs. 39%), $F_1(1, 23) = 27.70$, $p < .001$; $F_2(1, 35) = 13.00$, $p < .002$. Pairwise comparisons ($\alpha = .05$) with both subjects and items as random factors showed that as expected,

Table 3
Experiment 1, grammatical judgment accuracy

	Ferreira & Henderson	Van Dyke & Lewis
Short	62% (*SE* = .05)	85% (*SE* = .03)
Long/complex	26% (*SE* = .06)	38% (*SE* = .06)
Long/non-complex	29% (*SE* = .05)	53% (*SE* = .06)

subjects were more accurate for short region sentences (74%) than for non-complex long sentences (41%), with complex long sentences being the most difficult (32%). In addition, the difference between the two long sentence types was also significant for subject analyses, and marginally for items analyses ($p < .06$). No interactions were significant.

Considering just the F&H sentences, we found a main effect of region type, $F_1(2, 46) = 35.88$, $p < .001$; $F_2(2, 70) = 29.71$, $p < .001$. Subjects judged sentences with short ambiguous regions 62% accurately, while accuracy for long/complex and long/non-complex sentences was 26 and 29%, respectively. The difference between the short region and either of the two long regions was significant in pairwise t tests ($\alpha = .01$) in both subjects and items analyses. The difference between the two long regions, however, was not significant.

For the V&L sentences, the main effect of region type was also significant, $F_1(2, 46) = 31.56$, $p < .001$; $F_2(2, 70) = 39.74$, $p < .001$. Accuracy for short region sentences was 85%, while long, non-complex region (low interference) sentences were 53% and long, complex region (high interference) sentences was 38%. All pairwise comparisons among these three were significant ($\alpha = .03$). Short regions were easier than long, non-complex regions, and they were also significantly easier than long, complex regions. The difference between the two long regions was also significant, with the non-complex region being easier than the complex region.

Accuracy on grammatical filler sentences was 95%; accuracy on ungrammatical filler sentences was 93%.

Discussion

Our results using Ferreira and Henderson's (1991b) sentences replicated the accuracy measures they report almost exactly. For their early closure sentences, they reported accuracy rates of 63% for short sentences, 35% for long/complex sentences, and 34% for long/non-complex sentences. Like them, we found no effect of syntactic complexity in the ambiguous regions of their sentences, and only an effect of region length (i.e., distance from the head of the ambiguous region to the disambiguating word). Contrary to Ferreira and Henderson (1991b), however, this experiment demonstrated an effect of syntactic complexity of the ambiguous region with our sentences. Highly interfering sentences were significantly more difficult for subjects to recognize as grammatical than low interference sentences of the same length. Moreover, these sentences were controlled for head distance, the factor which Ferreira and Henderson suggest determines reanalysis difficulty. These results suggest that syntactic complexity, defined by our theory of retrieval interference, does in fact affect reanalysis success.

In addition, if we compare either of our long sentences with the short sentences, we observed an effect of

head distance predicted by Ferreira and Henderson (1991b) for both high interference and low interference sentences. It should be noted, however, that the low interference sentences provide the more appropriate estimate of the head distance effect because they are less confounded with interfering structure effects. We return to this observation in Experiment 2 below. The difference observed here supports the Ferreira and Henderson (1991a, 1991b) claim that head distance affects reanalysis in addition to syntactic structure.

Lastly, our results suggest that our sentences are easier for subjects than the F&H sentences. At first glance, this result is surprising because the ambiguous regions of the F&H sentences are shorter overall, and their long/non-complex sentence with the prepositional phrase in the ambiguous region has a much simpler syntactic construction than the interfering long/non-complex sentences with their embedded subject relative clauses. However, following Sturt et al. (1999), the difference can be explained in structural terms. The Ferreira and Henderson sentences contain NP/Z ambiguities, which are predicted to be more difficult than the NP/S ambiguities in the V&L sentences because of the type of structural change required. Thus, our results provide additional support for the effect of structural change on reanalysis, and we will provide a retrieval interference account of this effect in the General Discussion.

In addition, these results provide evidence for a second type of structural effect that is not dependent on the *type* of structural change, in that the high interference sentences were more difficult than the low interference sentences. Both sentences require a structural change that is predicted to be easy according to models which attribute reanalysis costs to function B (retracting commitments) (e.g., Gorrell, 1995; Sturt & Crocker, 1998; Sturt & Crocker, 1996, 1997; Weinberg, 1993). Consequently, these models must be modified to account for the structural effects observed here. Experiments 2–4 explore the interference effect in more detail, with goal of understanding the locus of this structural effect on attachment (function D) and repair (functions A–C).

Experiment two

In order to understand whether the structural effect observed in Experiment 1 is a property of attachment or repair, in this experiment we compared ambiguous and unambiguous versions of the V&L sentences used in Experiment 1. We suggested above that processing difficulties associated with highly interfering structures result from failure to retrieve items required for the reanalysis, a process inherent to the way the retrieval parser (i.e., attachment mechanism) works. Consequently, if our interference account is correct, then we

should see the interference effect with both ambiguous and unambiguous sentences. If, however, it is a factor relevant to the repair operation, then the interference effect should appear with ambiguous sentences only.

Method

Subjects

Thirty-six subjects from the University of Pittsburgh participated in the experiment for partial credit. All subjects were native speakers of American English.

Materials

One-half of the materials were the ambiguous V&L sentences used in Experiment 1 (see Table 1). The other half of the materials were non-ambiguous versions of these same sentences, constructed by adding the word *that* after the matrix verb of the sentence.

Procedure

The procedure was the same as in Experiment 1.

Design and analysis

The experiment was analyzed as a 2×3 repeated measures ANOVA with ambiguity (ambiguous, unambiguous) and region type (short, high interference, low interference) as factors. Separate ANOVAs were done on ambiguous and unambiguous sentences to determine pairwise effects within each sentence type.

Results

Table 4 shows the mean accuracy scores for the three conditions of each sentence type. We found main effects of both sentence type and region type. Overall, ambiguous sentences were judged correctly 61% of the time, while unambiguous sentences were judged accurately 85% of the time, $F_1(1, 35) = 96.51$, $p < .001$; $F_2(1, 35) = 59.87$, $p < .001$. For the region variable overall, subjects were 86% accurate on the short region sentences, 78% accurate on the low interference sentences, and 56% accurate for high interference sentences, $F_1(2, 70) = 49.17$, $p < .001$; $F_2(2, 70) = 38.45$, $p < .001$. Pairwise comparisons ($\alpha = .01$) indicate that all differences are significant. The interaction was not significant.

For ambiguous sentences alone, the region variable was significant, $F_1(2, 70) = 24.46$, $p < .001$; $F_2(2, 70) =$

19.87, $p < .001$. All pairwise differences were significant ($\alpha = .01$). Short region sentences were judged grammatical 77% of the time as compared with low interference sentences, which were judged grammatical 63% of the time. High interference structures were judged grammatical 44% of the time, which is different from both the short sentences and the low interference sentences.

For the unambiguous sentences only, the region effect was also significant, $F_1(2, 70) = 28.29$, $p < .001$; $F_2(2, 70) = 25.32$, $p < .001$. The short region sentences were judged accurately 94% of the time and low interference sentences were judged accurately just as often, 93%. The high interference sentences were accurately judged 69% of the time and this rate was significantly different from either the short sentences or the low interference sentences ($\alpha = .001$).

Recall that the unambiguous sentences provide an estimate of the cost associated with making the correct attachment for each sentence type. The ambiguous sentences will include this cost and all additional cost associated with the repair itself. Thus, in order to estimate costs of repair alone, attachment costs are factored out by subtracting accuracy for ambiguous sentences from that for unambiguous sentences. For interfering sentences, the repair cost for the short region was 17%, $t_1(35) = -4.48$, $p < .001$; $t_2(35) = -3.79$, $p < .002$; for high interference regions it was 25%, $t_1(35) = -5.01$, $p < .001$; $t_2(35) = -4.88$, $p < .001$; and for low interference regions it was 30%, $t_1(35) = -7.68$, $p < .001$; $t_2(35) = -6.36$, $p < .001$. Only the difference between the repair costs for the short sentences and low interference sentences was significant, $F_1(2, 35) = 4.58$, $p < .04$; $F_2(2, 35) = 7.81$, $p < .01$.

These materials afford a test of the two factors suggested by our previous results as contributing to the garden path effect: head distance and structural interference. The head distance effect can be estimated by comparing our sentences with a short ambiguous region, in which the disambiguating region follows the ambiguous region immediately, to the long low interference sentences, in which the long ambiguous region necessitates a much later disambiguating region, just as in the Ferreira and Henderson (1991a, 1991b) materials that yielded their original result (see Experiment 1). Similarly, the structural interference effect can be estimated by comparing the long low interference sentences and the long high interference sentences, since they are matched for position of the head as well as length of the ambiguous phrase. In addition, the effects of these variables on attachment (Function D) and repair (Functions A–C) operations can be partialed out by looking at the effect sizes for *unambiguous* conditions alone and on the *garden path effects*, respectively. Thus, the head distance effect on *attachment* is estimated by comparing the short and the long, low interference *unambiguous* sentences, while the interference effect on attachment is estimated by

Table 4
Experiment 2, grammatical judgment accuracy

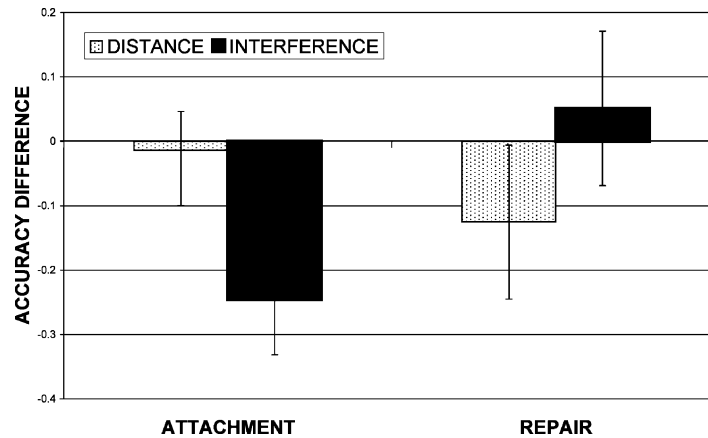|	Ambiguous	Unambiguous
Short	77% (SE = .04)	94% (SE = .02)
Interfering	44% (SE = .04)	69% (SE = .04)
Non-interfering	63% (SE = .04)	93% (SE = .02)

Fig. 6. Interaction of distance and interfering structure of the attachment and reanalysis functions in accuracy judgments to sentences in Experiment 2.

comparing the accuracy rates for the *unambiguous* sentences in the two long conditions.

Effects of head position on *repair* are determined by comparing the *garden path effects* for the short and low interference sentences, and interference effects on repair are determined by comparing the *garden path effects* for the two long sentences. These cost estimates have been plotted for attachment and repair operations in Fig. 6. A 2 (cost) $\times$ 2 (operation) ANOVA suggests that the main effects of operation and cost were not significant; however, their interaction was, $F_1(1, 35) = 9.29$, $p < .005$, $F_2(1, 35) = 7.71$, $p < .01$. We also conducted t tests to determine whether these estimates were different from zero. Costs associated with interference were significant only for attachment operations, $t_1(35) = -5.053$, $p < .001$; $t_2(35) = -5.808$, $p < .001$. Costs associated with distance were significant only for reanalysis operations, $t_1(35) = -2.795$, $p < .05$, $t_2(35) = -2.14$, $p < .05$. Thus, it appears that interference has its greatest effect for attachment, while costs associated with head position are more important for reanalysis operations.

Accuracy on grammatical fillers was 94%, and 95% on ungrammatical filler sentences.

Discussion

Experiment 2 replicated the structure effect found in Experiment 1. These data suggest that sentences with high interference ambiguous regions were more difficult to repair than those with low interference ambiguous regions. This finding was in addition to a replication of the Ferreira and Henderson (1991b) head position effect for both high interference and low interference sentences. The addition of unambiguous controls in the current experiment allowed us to separate costs related to attachment and costs related to repair. The results

indicate that the head distance effect is a property to which the repair processes (Functions A–C) are sensitive, and not the initial attachment processes (Function D). This is illustrated clearly by the significant interaction in Fig. 6, where attachment costs (i.e., those due to the parser's operation) have almost no component due to distance. Conversely, while Experiment 1 suggests that interference is a structural factor that does affect the reanalysis process as a whole, the results of this experiment show that its effect is primarily on the function of the process that makes the correct attachment and not those functions related to repairing the incorrect parse.

Experiment three

One potential problem with the results of Experiment 2 is that roughly 25% of the high interference materials in both the unambiguous and ambiguous conditions contained a second ambiguity in the embedded clause. For example, in sentence (16) below, the embedded verb *revealed* could take either an NP object or a clausal object.

(16) The executive assistant forgot the letter which revealed the mayor was responsible was waiting for a signature.

In leaving out a *that* in the embedded clause, this created a possibility that an additional reanalysis was necessary, and while this was not a problem for low interference sentences, it may have influenced processing on the high interference sentences in Experiment 2, causing them to be more difficult. To correct for this, Experiment 3 is a replication of Experiment 2, with a *that* inserted into the embedded clause in all the high

interference sentences for both the ambiguous and unambiguous conditions.[4]

Method

Subjects

Thirty-six subjects from the University of Pittsburgh participated in the experiment for partial credit. All subjects were native speakers of American English.

Materials

The same materials as Experiment 2 were used, except that in the high interference conditions the word *that* was added to the most-embedded clause for both ambiguous and unambiguous sentences.

Content norming

In addition to the double ambiguity in some of the materials, another possible objection to our interpretation of the interference effects demonstrated in Experiment 2 is that there are content differences between the two sentences. In order to determine whether such differences would cause subjects to respond differently to our high and low interference sentences, we conducted a content norming experiment prior to carrying out Experiment 3. In this experiment, we altered the syntactic structure of the sentences, while keeping the content identical to that found in the experimental sentences for Experiment 3. This was done by transforming a low interference sentence such as (17a) into a content match sentence as in (18a). Similarly, a high interference sentence such as (17b) was transformed into a content match sentence as in (18b).

(17a) The secretary forgot the student who knew that the exam was important was standing in the hallway.

(17b) The secretary forgot the student who was waiting for the exam was standing in the hallway.

(18a) The student knew that the exam was important, and the secretary forgot that he was standing in the hallway (.73).

(18b) The student was waiting for the exam, and the secretary forgot that he was standing in the hallway (.79).

---

[4] Note that this created a small length confound, since the high interference sentences were now one word longer in the ambiguous region. While we did not address this confound in the current experiment, Experiment 4 does address it by regressing length out of the reading times reported there. Since the results there were similar to those reported in the current experiment, we do not believe this confound presents a problem for interpreting the results of Experiment 3. Moreover, the result of this manipulation was to make the high interference sentences easier, rather than more difficult, as might be predicted by their longer length.

As with the previous experiments, subjects were asked to judge the grammaticality of these sentences. Twenty-four subjects participated, yielding the accuracy scores shown in parenthesis after each sentence in (18a and 18b). While judgments appear quite low for these sentences, we suggest that this may be due to uncertainty about the referent for the pronoun in the second clause. Statistically, the difference between the two conditions was not significant, $t_1(23) = 1.40$, $p < .17$; $t_2(23) = 1.25$, $p < .22$, suggesting no content-specific effects on grammatical judgments.

Procedure

The procedure was the same as in Experiment 2.

Design and analysis

The design and analysis was identical to Experiment 2.

Results

Table 5 shows the mean accuracy scores for the three conditions of each sentence type. Replicating the results from Experiment 2, we found main effects of both sentence type and region type. Overall, ambiguous sentences were judged grammatical 58% of the time, while unambiguous sentences were judged accurately 90% of the time, $F_1(1, 35) = 117.35$, $p < .001$; $F_2(1, 35) = 92.42$, $p < .001$. For the region variable overall, subjects were 86% accurate on the short region sentences, 75% accurate on the low interference sentences, and 61% accurate for high interference sentences, $F_1(2, 70) = 33.00$, $p < .001$; $F_2(2, 70) = 25.61$, $p < .001$. The interaction of these factors was significant, $F_1(2, 70) = 7.19$, $p < .002$; $F_2(2, 70) = 6.05$, $p < .005$. Separate analyses for the ambiguous and unambiguous sentences indicate that the interaction effect is due to a difference between the short and the low interference long sentences in the ambiguous condition, but not in the unambiguous conditions.

For ambiguous sentences alone the region variable was significant, $F_1(2, 70) = 21.39$, $p < .001$; $F_2(2, 70) = 17.78$, $p < .001$. Short region sentences were judged grammatical 76% of the time as compared with low interference sentences, which were judged grammatical 56% of the time, ($p < .001$ by subjects and items). High interference structures were judged grammatical 43% of the time, which was significantly less than the short sentences

Table 5
Experiment 3, grammatical accuracy

	Ambiguous	Unambiguous
Short	76% (*SE* = .04)	95% (*SE* = .01)
Interfering	43% (*SE* = .05)	80% (*SE* = .05)
Non-interfering	56% (*SE* = .04)	94% (*SE* = .02)

($\alpha = .001$), but marginally less than the low interference sentences ($p < .06$ by subjects, $p < .02$ by items).

For the unambiguous sentences only, the region effect was also significant, $F_1(2, 70) = 19.57$, $p < .001$; $F_2(2, 70) = 15.54$, $p < .001$. The short region sentences were judged accurately 95% of the time, while the low interference sentences were judged correctly 94% of the time. The high interference sentences were judged accurately 80% of the time and pairwise comparisons ($\alpha = .01$) indicate that this rate was significantly different from either of the other two conditions.

As in Experiment 2, we estimated the garden path effect for each sentence type by comparing costs associated with making the correct attachment, as estimated by unambiguous sentences and those associated with the repair, as estimated by the difference between the ambiguous and unambiguous versions (i.e., the garden path effect) of the sentences for each region type. In this experiment, the garden path effect for the short sentences was 19%, $t_1(35) = -5.24$, $p < .001$; $t_2(35) = -4.90$, $p < .001$; for the high interference sentences 38%, $t_1(35) = -7.11$, $p < .001$; $t_2(35) = -7.77$, $p < .001$; and 37%, $t_1(35) = -9.81$, $p < .001$; $t_2(35) = -7.35$, $p < .001$ for the low interference sentences. A comparison of these effects showed a significant effect of region type, $F_1(2, 70) = 7.19$, $p < .002$; $F_2(2, 70) = 6.05$, $p < .005$, due to the difference between the short sentences and either of the two longer sentence conditions.

As in Experiment 2, we calculated estimates for the distance and interference effects on attachment and repair operations, and these estimates are plotted in Fig. 7. Recall that the comparison of the short and low interference sentences provide an estimate of the costs due to distance, while the difference between the high interfer-ence and low interference sentences provide an estimate of costs associated with attachment. A 2 (cost) × 2 (operation) ANOVA revealed a significant interaction, $F_1(1, 35) = 7.88$, $p < .01$, $F_2(1, 35) = 6.92$, $p < .01$, but no significant main effects. t tests were used to compare these estimates with zero, or no cost for attachment or distance. The interference effect was significantly differ-ent from zero only for attachment costs, $t_1(35) = -4.78$, $p < .001$; $t_2(35) = -4.21$, $p < .001$, and the distance ef-fect was significant only for repair, $t_1(35) = -3.73$, $p < .002$; $t_2(35) = -2.91$, $p < .01$.

Accuracy on grammatical fillers was 95% and 93% on ungrammatical filler sentences.

Discussion

Recall that Experiment 3 was designed to replicate the results of Experiment 2, while removing a potential source of error in the high interference conditions which may have contributed to the difficulty in those condi-tions observed in Experiment 2. As expected, the addi-tion of a disambiguating *that* reduced difficulty in the unambiguous high interference condition in the current experiment; however, this addition had little effect on the error rate for the ambiguous sentences. This suggests that the initial ambiguity of the object of the matrix verb was the main contributor of processing difficulty in these sentences, with little additional difficulty caused by ambiguities in the embedded clause, perhaps because the error rate was already reflecting a floor effect. Even with this change between the experiments, the results of Experiment 3 are consistent with those in Experiment 2. In both experiments the short and the long low inter-ference sentences were not different from each other in the unambiguous conditions, while they did differ in the
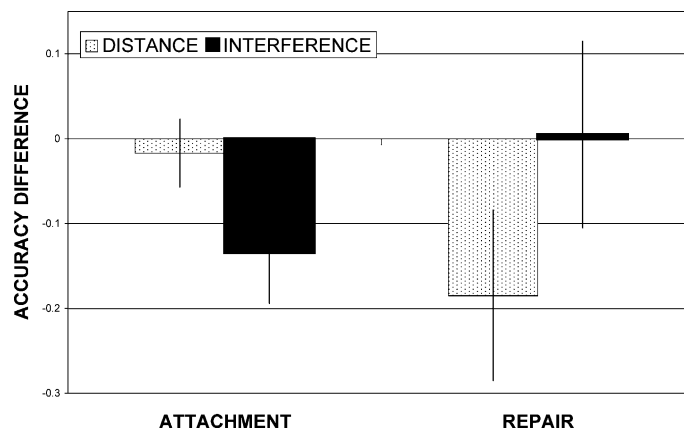

Fig. 7. Interaction of distance and interfering structure of the attachment and reanalysis functions in accuracy judgments to sentences in Experiment 3.

ambiguous conditions.[5] This would be expected if the greater distance between the subject of the embedded clause and its verb had no effect on attaching the subject into the existing parse tree, as was predicted by the operation x cost interaction observed in Experiment 2. This is also evident in the fact that the garden path effects for both the high and low interfering conditions did not differ (37% and 38%, respectively). Despite the difference between the two interfering sentences in the ambiguous conditions, we see from comparing them with their unambiguous controls that the difference is due to the attachment difficulties, and not due to the reanalysis itself. Consequently, Experiment 3 provides a clear replication of the differential effect of interference on attachment and head distance on repair.

As noted earlier, these data suggest that structure, when measured in terms of retrieval interference, does affect reanalysis. This effect is by virtue of the fact that reanalysis must recruit the mechanisms of the parser in order to succeed. There are important implications for the processes specific to reanalysis, in that the amount of interference generated by the material between the ambiguity and the disambiguation point limits whether the repair will be successful. For example, consider the structure built for the ambiguous sentence in (17a). In order for the reanalysis to be successful, the parser must be able to retrieve *forgot* and attach the clause headed by *was standing* as its complement. Yet because of the interference from the intervening structure, this item is difficult to retrieve and will therefore be unavailable to participate in the structural changes required by the reanalysis. In contrast, the low interference sentence in (17b), which is matched with (17a) for length of ambiguous region and head distance to the disambiguating word, does not suffer from this problem, as there is no interfering sentential complement verb that would make *forgot* difficult to retrieve. Consequently, reanalysis succeeds at a much greater rate than for high interference sentences, (56% vs. 43%), because the retrieval of the necessary NP permits creation of the clausal complement. While the success rate for these low interference sentences is still not as great as with the short ambiguous region sentences (56% vs. 76%), this is due to costs associated with the repair operation alone, as can be seen from the statistically similar error rates in the unambiguous versions of these sentences,

which do not require reanalysis. The data presented here suggest that this does not result from structural factors that make required constituents unavailable, but rather from the added distance separating the ambiguous head from the disambiguating word. It might be argued, however, that the differential effect of distance on repair, but not attachment, is due to a ceiling effect for the short and low-interfering long sentences.[6] Since the low-interfering sentences are accurately judged to be grammatical at the high rate of 94%, there is little room for the decreased distance between the ambiguous NP and the disambiguating item in the short sentences to make a difference. Consequently, any effect of distance on the unambiguous items may be undetectable in the current experiment. Experiment four addresses this concern by replicating the current experiment in an on-line reading task.

Experiment four

In order to support comparisons with the original Ferreira and Henderson (1991b) data regarding the role of syntax and head position on reanalysis, Experiments 1–3 used grammaticality judgments as the dependent measure. Since this measure provides only an off-line assessment of processing difficulty, it is desirable to explore the role of interference and head position with online reading measures. To that end, Experiment 4 uses the same materials as those from Experiment 3 in a reading comprehension task that allows for the examination of reading times as the subject moves through the sentence. In order to verify the dissociation between interference and distance on attachment and reanalysis costs observed in Experiments 2 and 3, we perform identical analyses on the reading time data collected here.

Reading time data also allows us to compare predictions regarding the locus of effects according to the cue-based retrieval account and a storage account such as the Dependency Locality Theory (DLT), which makes specific predictions about where storage effects occur during processing of complex sentences (Gibson, 2000). As we have seen, Ferreira and Henderson (1991b) have focused on the number of words as the increment for their distance-based theory, but Gibson and colleagues (Gibson, 1998, 2000; Gibson & Thomas, 1999; Warren & Gibson, 2002) present evidence arguing for an alternative metric for accounting for the difficulty of long-distance attachments: the number of discourse referents that the parser must create and maintain in memory before making the correct attachment.

According to this model, parsing costs are determined by three types of resource expenditure: (1) those associated with creating new discourse referents; (2) those associated with attaching (or integrating) gram-

---

[5] This created a significant interaction of ambiguity by region type in Experiment 3, but the same interaction did not reach significance in Experiment 2, because of the increased error variance in the unambiguous sentences caused by the spurious ambiguity in the materials. Our account does not depend on the significance of this particular interaction, however, because the crucial effects for our argument are the difference between the low and high interfering unambiguous conditions, and the lack of a difference between the garden path effects for these two conditions, both of which were present in both experiments.

[6] Thanks go to Brian McElree for pointing this out.

matical heads and dependents; and (3) those associated with storing incomplete structural dependencies (i.e., syntactic predictions). These costs are illustrated in Table 6 for the low interference sentence in (19a).

(19a) The older boy understood that the girl who was swimming near the dock was paranoid about dying.

(19b) The older boy understood that the girl who said that the townspeople were dangerous was paranoid about dying.

The DLT assigns 1 energy unit (EU) for each new contentful discourse referent, which in the case of (19a) includes nouns, verbs, and prepositions.[7] This is illustrated in the first row of Table 6. As already noted, attachment costs are incremented by the number of intervening discourse referents between the head and dependent. This is illustrated in the second row of Table 6. Notably, the only long distance attachment in this sentence is between the final VP *was paranoid* and the NP *the girl*. The cost of 4 EUs refers to the referents *swimming*, *near*, *dock*, and *paranoid*, which must be crossed to make this attachment. Storage costs arise because the parser is thought to make predictions about the syntactic constituents expected in the sentence. For example, when the parser sees the first word *the*, it makes a prediction for both a subject and verb for the upcoming sentence, producing a storage cost of 2. When the subject *boy* occurs, one of these predictions is realized, and the storage cost is reduced to 1. Similarly, when the matrix verb *understood* occurs, the second initial prediction is realized, and storage cost is reduced to zero. When the complementizer *that* occurs, it is incremented back to 2, since *that* signals an upcoming subject and verb. When *girl* occurs, this cost is decremented because the subject prediction is realized. The cost is incremented again at *who* because now not only must the parser store the prediction of the verb to go with *girl*, but also for a subject and verb for the relative clause signaled by the relative pronoun. When *swimming* occurs, both relative clause predictions are satisfied, but the verb prediction for *girl* is still outstanding. When the proposition *near* occurs, a prediction for an object for the preposition must be stored in addition to the verb prediction, resulting in a storage cost of 2. This is again

---

[7] Gibson (2000) does not address sentences containing prepositions, but Gibson (1998) does suggest that prepositions are counted as new discourse referents. In no presentations of the DLT/SPLT does Gibson address compound verbal predicates such as was paranoid in the sentences considered here. Since the predicate paranoid provides the semantic content of the verb phrase, we assign all VP-related costs to this word and not to the copular verb was. The arguments presented here regarding the DLT as compared to the Interference theory do not depend on this treatment however, since a different assignment would still lead to the same prediction regarding the locus of highest processing cost according to the DLT.

Table 6
Word by word predictions of the discourse locality theory for the low interference sentence in (19a)

Cost type	Input word															
	The	boy	understood	that	the	girl	who	was	swimming	near	the	dock	was	paranoid	about	dying
New discourse referent	0	1	1	0	0	1	0	0	1	1	0	1	0	1	1	1
Attachment	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0
Storage	2	1	1	2	2	1	3	3	1	2	2	1	1	0	1	0
Total	2	2	1	2	2	2	3	3	2	3	2	2	1	5	2	1

Table 7
Word by word predictions of the discourse locality theory for the high interference sentence in (19b)

Cost type	Input word															
	The	boy	understood	that	the	girl	who	said	the	townspeople	were	dangerous	was	paranoid	about	dying
New discourse referent	0	1	1	0	0	1	0	1	0	1	0	1	0	1	1	1
Attachment	0	0	0	0	0	0	0	0	0	0	0	1	0	4	0	0
Storage	2	1	0	2	2	1	3	3	3	2	2	1	1	0	1	0
Total	2	2	1	2	2	2	3	4	3	3	2	3	1	5	2	1

decremented when the NP *dock* satisfies the prepositional prediction, leaving only the outstanding verbal prediction, which is finally realized when *paranoid* occurs.

In like fashion, Table 7 presents the DLT costs for the high interference sentence in (19b). Notably, both the low and the high interfering sentences contain the same number of new discourse referents, and require a single non-local integration predicted to cost (4 EU) at the same point in processing (i.e., at the VP head *paranoid*.) Thus, it is only the storage component of the DLT that can account for the empirical evidence that high interfering sentences like (19b) are more difficult than low interfering sentences like (19a). The shaded regions in Tables 6 and 7 indicate that the locus of the predicted difference according to the DLT occurs in the region of the embedded complement clause in the high interference condition. This is because the interfering sentence has two additional unrealized syntactic predictions when the verb *said* occurs, incrementing storage cost beyond the level occurring in the analogous region in the low interfering sentence. Critically, the DLT does *not* predict reading time differences in the area of the final VP *was paranoid*, since both the low and the high interfering sentences require a long distance attachment costing 5 EUs.

This is in direct contrast to the predictions of the Interference theory. Recall that retrieval interference is measured by the extent to which retrieval cues provided by the head of the phrase, the verb in the current case, unambiguously identify the correct dependant from previously encountered candidates. Since all attachments in sentences (19a and 19b) are local until the final long distance integration of *girl* and *was paranoid*, no significant processing difference between the high and low interference sentences is predicted until the final VP occurs, at which point a main effect is expected in which the high interference sentences will be more difficult than the low interference sentences. Thus, reading times will provide evidence for distinguishing a retrieval theory such as the Interference theory, which predicts the processing difference to appear in the disambiguating region, with no effects in the ambiguous region, and a storage account such as the DLT, which predicts the difference to occur in the ambiguous region itself, with no effect in the disambiguating region predicted.

Method

Subjects

Thirty-seven subjects from the University of Pittsburgh participated in the experiment for course credit. All subjects were native speakers of English, whose vision was corrected to normal. This odd number of subjects was due to a scheduling error resulting in one extra subject than planned. We included the subjects' data in the analysis since all comparisons were within subject comparisons, and no effects of list were observed.

Materials

The exact materials used from Experiment 3 were used here.

Procedure

As in Experiments 1–3, subjects read each sentence one word at a time, presented in a non-cumulative moving-window format. At the end of each sentence, subjects were presented with a yes/no comprehension question, which, in the case of filler sentences, queried either the subject or the object of matrix verb in the sentence they had just read. As in Experiments 1–3, subjects responded using "yes" and "no" keys on the keyboard, designated as the "1" and "2" keys. For all the experimental sentences, the comprehension question pertained to the long distance attachment required for integrating the final verb into the sentence. For example, the comprehension question for all the conditions of the V&L sentences shown in Table 7 would be "*Was it the man who was paranoid?*"

Reading time analysis

Sentences were partitioned into five regions of interest, as illustrated by slashes in the sentences in Table 8.

The first region contained the subject NP and the matrix verb in all conditions; however, we did not include the first two words of the sentence because these contained longer reading times associated with the beginning of the sentence (indicated by the double slashes). The second region contained the length manipulation, and hence was not defined for the short sentence. For the low interference conditions, it contained the relative clause *who was waiting for the exam*, and for the high interference conditions it contained the clause *who knew that the exam was important*. Note that this is the region

in which we expect to find evidence of any effect due to additional storage demands for the high interfering sentences. Region three was the critical region for re-analysis, as it contained the disambiguating VP, *was standing*, in all conditions, and it is this region which will show effects due to retrieval interference. Region four was a final spillover region, containing up to two words, which in this example are *in the*. The final word, separated from region 5 by double slashes, was not considered in the analyses because the reading times for this word are assumed to include sentence-final wrap up effects (cf. Just & Carpenter, 1980; King & Just, 1991; Mitchell & Green, 1978; Rayner, Kambe, & Duffy, 2000 for similar technique).

Data analyses were conducted on residual reading times, which enabled us to remove variance associated with subject-specific reading rates and differences in region lengths. A regression equation for each subject was calculated using word length and word position as predictors of the reading time for each word. This procedure has been advocated by Ferreira and Clifton (1986) in order to correct for the fact that a word's reading rate is a function of the number of characters (Rayner, 1977; Rayner, Sereno, & Raney, 1996) but that this function does not have a zero-intercept due in part to costs associated with the button-press in a self-paced reading paradigm. The regression equation provides an estimate of costs associated with this button press, together with a multiplier for predicting the reading time for a given word based on its length. In addition to the word length predictor, we included a word position predictor in our regression analyses to correct for a possible gradual decrease in reading speed as subjects move through the sentence. While a number of researchers have advanced the explanation to account for aspects of their data

Table 8
Sample items, Experiment 4, marked for reading time regions

Ambiguous versions
Short region
The executive // assistant forgot the student / was standing / in the // hallway.

High interference region
The executive // assistant forgot the student / who knew that the exam was important/ was standing / in the // hallway.

Low interference region
The executive // assistant forgot the student / who was waiting for the exam / was standing / in the // hallway.

Unambiguous versions (identical to above, but with added that following the matrix verb)
Short region
The executive // assistant forgot that the student / was standing / in the // hallway

High interference region
The executive // assistant forgot that the student / who knew that the exam was important/ was standing / in the // hallway

Low interference region
The executive // assistant forgot that the student / who was waiting for the exam / was standing / in the // hallway

---

Table 9
Coefficients for first-order linear regression equation: $y = B_0 + B_1x + B_2x$

	B_0 (constant)	B_1 (word position)	B_2 (word length)
Mean coefficient	343.08	−2.60	14.20
$SD =$	47.17	3.78	8.01
$t(36) =$	44.24	−4.19	10.78
$p <$	.001	.001	.001

(e.g., Ferreira & Henderson, 1993; Gernsbacher & Faust, 1990; Sturt et al., 1999) none have attempted to statistically correct for the effect. In our analyses, such a correction is essential because we will be comparing a critical region occurring as the 7–9th words in our short sentences and as the 15–17th words in our longer sentences. Moreover, the regression equations for our individual subjects reveal that 28 of our 37 subjects had negative coefficients for the word position predictor ($M = -2.60$), suggesting that these subjects' reading time did speed up as they moved through the sentence. Table 9 presents the average coefficients for the predictors in each equation, as well as t tests indicating that they are all significantly different from zero.

The regression equations were calculated using data from fillers only, excluding any per word reading times that were less than 150 ms or greater than 2000 ms. The residual reading time for each word in the experimental sentences was then calculated by subtracting the predicted reading time from the observed time. Residual reading times for each region were then calculated by taking the aggregate of all words in the region of interest. Following Sturt et al. (2001) we trimmed outliers in each region by replacing the values that were either three times the interquartile range below the lower quartile or above the upper quartile for each condition with the appropriate cutoff value. These data trimming procedures affected 1.7% of the data.

Results

Comprehension questions

Table 10 presents mean accuracy scores for the comprehension questions in each condition. We observed a main effect of ambiguity that was significant for the subjects analysis, $F_1(1, 36) = 9.31$, $p < .005$; but not for the items analysis, $F_2(1, 36) = 2.90$, $p < .10$.

Table 10
Experiment 4, comprehension questions

	Ambiguous	Unambiguous
Short	91% ($SE = .02$)	93% ($SE = .02$)
Interfering	71% ($SE = .03$)	83% ($SE = .02$)
Non-interfering	87% ($SE = .02$)	87% ($SE = .02$)

As would be expected, this effect was due to subjects having more difficulty understanding the ambiguous sentences than the unambiguous ones. The main effect of region was also significant, $F_1(2, 72) = 18.02$, $p < .001$; $F_2(2, 70) = 10.68$, $p < .001$. This effect was due to the high interference condition being more difficult to understand than either the short or the low interference condition. The interaction between these factors was also significant in the subjects analysis, $F_1(2, 72) = 4.58$, $p < .02$; but only marginally so in the items analysis, $F_2(2, 70) = 2.97$, $p < .06$. This is due to the ambiguity effect being significant only for the high interference sentences, $t_1(36) = -3.62$, $p < .001$; $t_2, (36) = -2.35$, $p < .03$; all other $ts < 1$. In addition, a comparison of the ambiguity effect in each of the three sentence types revealed that the high interference condition was different from either the short condition, $t_1(36) = 2.64$, $p < 02$; $t_2(35) = 2.0$, $p < .05$ or the low interference condition, $t_1(36) = -2.44$, $p < .02$; $t_2(35) = 1.84$, $p < .08$ but that the short and low interference conditions were not different from each other, $t < 1$.

Reading times

Table 11 presents the residual reading times for each of the five regions of interest, together with untrimmed raw reading times. Note that because of the positive correlation between word position and reading time, the raw reading times are difficult to interpret, but are included here for completeness. We report statistics on residual analyses only, as analyses with raw scores yielded statistically identical results, except where noted. There were no significant effects in Region 1, all $Fs < 1.93$, $ps > .17$. An effect of ambiguity was observed in the Region 1 raw reading times, $F_1(1, 36) = 222.16$, $p < .001$; $F_2(1, 36) = 140.39$, $p < .001$ (see Table 8). This effect was likely due to the fact that in the unambiguous conditions, region 1 contained the extra word *that*. The effect was not observed in the analyses with residuals, since the length correction in the regression analyses provided a method of controlling for the extra word. Similarly, there were no significant effects in Region 2, the storage region, with all $Fs < 1.34$, $ps > .25$. An effect of interference type was observed in the Region 2 raw reading times, $F_1(1, 36) = 91.25$, $p < .001$; $F_2(1, 36) = 53.63$, $p < .001$ (see Table 8). This effect was likely due to the fact that the interfering conditions had one additional word in region 2. As in region 1, this effect was not observed in the analyses with residuals because of the length correction provided by the regression analyses.

The analyses for Region 3, the critical region containing the disambiguating verb phrase, suggest main effects of ambiguity and region type. The ambiguity effect was as expected, the ambiguous sentences were slower than the unambiguous sentences, $F_1(1, 36) = 13.66$, $p < .001$; $F_2(1, 35) = 13.59$, $p < .001$. As expected

Table 11
Experiment 4, trimmed residual reading times for each region, with untrimmed raw reading times in parenthesis (ms)

	Region 1 assistant forgot (that) the boy	Region 2 who knew that the exam was important/who was waiting for the exam	Region 3 was standing	Region 4 in the
Ambiguous				
Short	−15 (1601)		48 (938)	85 (674)
High Interference	24 (1659)	99 (2811)	161 (1044)	34 (573)
Low interference	−17 (1622)	171 (2500)	92 (939)	13 (555)
Unambiguous				
Short	−41 (1979)		50 (951)	61 (636)
High Interference	−45 (1952)	132 (2820)	62 (907)	50 (581)
Low interference	−6 (1998)	133 (2451)	6 (860)	20 (555)

from Experiment 3, region type was also significant, $F_1(2, 72) = 6.42$, $p < .005$; $F_2(2, 70) = 8.86$, $p < .001$, and the results followed the pattern found there. Pairwise comparisons indicated that the high interference sentences were slower than either the short ($p < .01$) or the low interference ($p < .005$) sentences. The difference between the short and the low interference sentences was not statistically reliable. The ambiguity × region interaction was also significant, $F_1(2, 72) = 6.62$, $p < .005$; $F_2(2, 70) = 3.44$, $p < .04$.

For the ambiguous sentences alone, the region effect was significant, $F_1(2, 72) = 8.42$, $p < .001$, $F_2(2, 70) = 7.38$, $p < .001$, and pairwise comparisons suggest that this is due to the high interference sentence being different from either the short sentence ($p < .001$) or the low interference sentence ($p < .01$). The difference between the short and the low interference sentences did not reach significance ($p = .10$).

For the unambiguous sentences, the region effect was significant in the subjects analyses, $F_1(2, 72) = 3.51$, $p < .04$, but not quite by items, $F_2(2, 70) = 2.82$, $p = .07$. Pairwise comparisons reveal that only the difference between the high interference and the low interference conditions was reliable for the unambiguous conditions ($p < .01$). The short condition was not quite different from the low interference condition ($p = .09$), nor from the high interference condition ($p = .59$).

Experiments 2 and 3 suggested that the differences between the ambiguous conditions could best be interpreted with respect to their unambiguous controls. Consequently, we calculated the garden path effect for each condition by comparing the ambiguous and the unambiguous sentences in that condition. The garden path effect for the short sentences was 2.28 ms (residual), which was not a significant difference, $ts < 1$. The garden path effects for both the longer sentences was significant. For the high interference condition the effect was 99 ms (residual), $t_1(36) = 3.65$, $p < .001$; $t_2(36) = 2.98$, $p < .01$; and for the low interference condition it was 86 ms (residual), $t_1(36) = 3.57$, $p < .001$; $t_2(36) = 3.00$, $p < .01$.

Only the difference between the short condition and either of the two interference conditions was significant, $F_1(2, 72) = 6.62$, $p < .005$; $F_2(2, 72) = 3.44$, $p < .04$.

As in Experiment 3, we analyzed the separate effects of interference and distance on attachment and repair operators by comparing the relevant conditions within our design. Fig. 8 presents these results in graphic format. As with the grammaticality judgment data, we conducted t tests to verify whether these effect estimates were different from zero, or no cost. The interference effect on the attachment operator was significant, $t(36) = 2.82$, $p < .01$, but not for repair, $t < 1$. The distance effect on the repair operator was significant, $t(36) = 3.21$, $p < .005$, but not for attachment, $t(36) = -1.737$, $p < .09$. The cost estimate for the effect of interference on repair was not significant, and nor was the effect of distance on attachment. As in previous experiments, we submitted these effect estimates to a 2×2 ANOVA comparing operation (attachment/repair) and cost (distance/interference). As expected, we found a significant interaction in the subjects analysis, $F_1(1, 36) = 4.78$, $p < .04$; however, this effect did not quite reach significance in the items analysis, $F_2(1, 35) = 3.08$, $p < .09$. The main effects of operation type (attachment/repair) and cost (distance/interference) were not significant in this analysis.

Finally, Region 4, the spillover region, revealed only a significant effect of region type, $F_1(2, 72) = 10.73$, $p < .001$; $F_2(2, 70) = 7.14$, $p < .005$. Pairwise comparisons suggest that all conditions differed from one another, ($ps < .05$), with the short condition being the slowest, followed by the high interference condition and then the low interference condition. The interaction of region type and ambiguity was not significant.

Discussion

The reading time results from Experiment 4 show the same pattern of effects observed in the two previous off-line experiments: interference has its primary
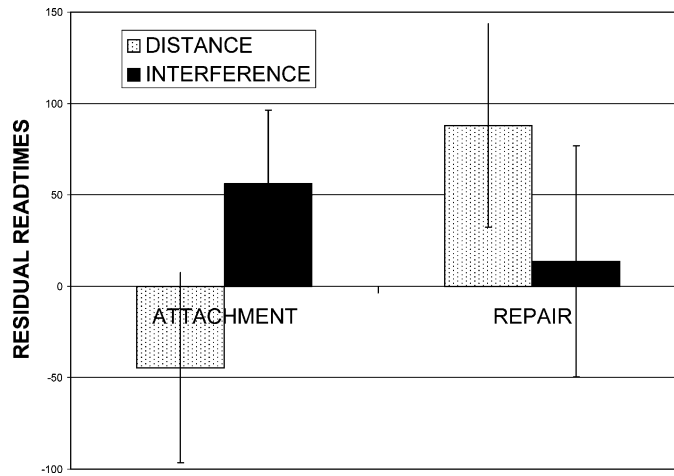
Fig. 8. Interaction of distance and interfering structure of the attachment and reanalysis functions in reading times at the disambiguating region in Experiment 4. Error bars indicate the 95% confidence intervals for whether the cost estimates are different from zero (or no cost).

effect on attachment operations, while head position, or distance between the ambiguous head and its disambiguator, has its primary effect on repair operations. Unlike in the offline data from Experiments 2 and 3, some evidence of distance effects on attachment and interference effects on repair was present in the online data, but the effects were not significantly different from zero.

Although the effects of structural complexity were clearly present in our data, we found no evidence of storage effects in the Region 2 reading times, and therefore no support for the predictions of a storage account such as the DLT of Gibson and colleagues (Gibson, 1998; Warren & Gibson, 2002) (cf. Van Dyke (2002) for further discussion). Rather, the results presented here favor the retrieval interference account of complexity effects, since the location of these effects coincides with the point of retrieval and not the point of maximum storage complexity in the sentence. It appears therefore that these structural effects are due to the processor's inability to identify the appropriate attachment sites specified by the retrieval cues of the final verb. Interfering sentences are difficult because the correct attachment cannot be made, even when no additional grammatical ambiguity is present. Although this is a property of cue-based retrieval parsing, and not reanalysis per se, it provides important constraints on whether the repair operation can successfully occur because a necessary component repair is the attachment of the disambiguating item. For example, consider the sentence in (20).

(20) The assistant forgot the letter was waiting for a signature.

A cue-based retrieval account of the reanalysis of this sentence proceeds as follows: when *forgot* is encoun-

tered, a lexical frame with an NP complement and one with a sentential complement are both retrieved. The occurrence of NP *the letter* causes it to be attached as the NP complement of *forgot*, allowing the alternative lexical frame to be discarded or ignored and begin decaying from memory. When the disambiguating verb *was* occurs, it provides a retrieval cue for an expected sentential complement. This cued retrieval causes the alternative lexical frame for *forgot* to be reactivated (and the alternative lexical frame deactivated via lateral inhibition) and the clause headed by *was* is attached as the complement it was expecting. Since *was* still has an empty subject role to fill, this prompts a second retrieval of the now detached NP *the letter* for its subject, and reanalysis is complete. Note, however, that if the ambiguous NP *the letter* were longer due to the inclusion of prepositional phrases, or relative clauses, this would delay the appearance of the disambiguating item and give the alternative lexical frame more time to decay, resulting in more difficult reanalyses (Ferreira & Henderson, 1991b; Warner & Glass, 1987).

Thus, the success of this reanalysis depends on the cued-retrieval of the alternative lexical frame for *forgot*, as this allows the disambiguating verb to be attached into the existing parse tree, which in turn allows the additional restructuring (i.e., re-attachment of *the letter* as subject of *was*) to occur as a natural consequence of the cue-based retrievals associated with a verb needing to fill its subject role. Data from the comprehension questions supports this account as well. We found that the ambiguity effect was significant only for the highly interfering conditions, with ambiguous versions comprehended only 71% of the time, reflecting a high failure rate of the reanalysis operation, over and beyond difficulties associated with making the correct attachment.

According to our account, this occurs because the highly interfering structure makes the re-structuring difficult because the retrieval cues from the verb cannot unambiguously identify the proper attachment site. In contrast, reanalysis occurred successfully in both the short and the low interference long sentences, confirming that structure of the intervening region in the low interference sentences did not affect restructuring. Interestingly, however, although the comprehension questions for the low interference long sentences were answered at a statistically similar rate as those of the short sentences, the reading times for this comparison suggest that it did take longer to arrive at the correct interpretation in the low interference conditions. Our account suggests that this is because the subject NP has had more time to decay during processing of the intervening region, requiring more time to reactivate the item and incorporate it into the parse tree. Critically, however, decay alone was not sufficient to prevent comprehension in these cases, because the necessary attachment was not hindered by ambiguous retrieval cues, as is the case with highly interfering structures. Thus, our results suggest that a comprehensive theory of sentence interpretation and reanalysis must include a role for structure effects as well as those associated with the decay of necessary lexical items.

General discussion

We have argued that syntactic attachments are the result of cue-based retrieval parsing, in which the head of a grammatical dependency provides cues to identify its dependents. When these cues are unambiguous, attachment occurs easily, as is the case with our low interference sentences. However, when the available cues do not disambiguate among competing candidates, attachment becomes costly because the correct dependent has a lower probability of being retrieved, as with our high interference sentences. This interference effect is a structural effect that arises because of the way the parsing mechanism operates, and consequently, the effects are observed for both unambiguous and ambiguous sentences, as both rely on cue-based retrievals in order to build (or revise) structure.

In addition, our data support Ferreira and Henderson's (1991b) suggestion that reanalysis is affected by distance between the points of ambiguity and disambiguation because the items required for the correct analysis are subject to decay. While studies attempting to provide online evidence of the head position effect have had mixed success at replicating Ferreira and Henderson's original result in reading times, Experiment 4 suggested that this may be due to analyses which failed to account for a gradual speed-up of per character reading times as subjects progress through a

sentence. Having partialed out these effects, we found clear indication of the head position effect, but only for ambiguous sentences, suggesting that the effect is due to the difficulty of retrieving discarded lexical frames necessary for the reanalysis. In contrast, no such retrieval is required for unambiguous sentences, since the globally correct lexical frame has not decayed, resulting in a less effortful retrieval of the attachment site for integrating the final verb phrase into the existing sentence, as evidenced by the lack of a distance effect for these conditions.

This evidence of the head position effect is important for distinguishing accounts of reanalysis. We have already seen that the Sturt et al. (1999) model, which focuses on costs associated with the *type* of structural change (Function B), requires additional modification to accommodate these results since they occur in structures which their model predicts to be easy to reanalyze. Similarly, the Fodor and Inoue (1994, 1998) account which focuses costs on Function A also requires modification, despite the fact that like the cue-based retrieval model, reanalysis is initiated by an attachment mechanism (Function D). In their model an *Attach Anyway* mechanism uses the same information available to the first pass parser's *Attach* mechanism, but loosens grammatical constraints to allow attaching an item into the existing parse tree at the point which least violates the grammar (see Lewis (1998) for a similar approach). According to their account, the momentary inconsistency created by the fact that *strike* in sentence (21) cannot take both an NP object and a sentential complement causes an *Adjust* operator to clean up the parse by altering less recent attachments, which as illustrated in Fig. 9, means breaking the attachment of *the dog* as the object of the verb *strike*. This leaves the NP free to serve as the subject of the embedded clause headed by *kicks*, and an *Adjust* operator now creates this dependency within the existing dependency created by *Attach Anyway* (i.e., between *strikes* and *kicks*).
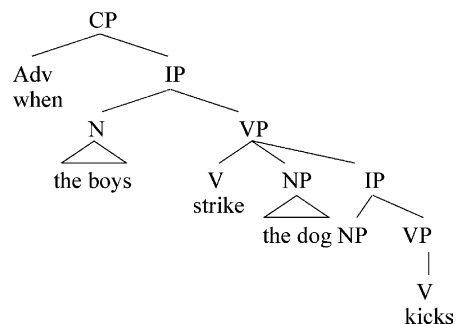
(21) When the boy strikes the dog kicks.

Fig. 9. Double attachment created by Fodor and Inoue's (1994, 1998) *Attach Anyway* operator as a trigger for reanalysis.

The Fodor and Inoue account differs from that of the cue-based retrieval parser because they predict no additional difficulty for reanalyzing a sentence such as (22a) as compared with (22b), which has a much shorter ambiguous region, because *Adjust* requires only a grammatical dependency between *forgot* and *was* to complete the reanalysis.

(22a) The assistant forgot the letter which had fallen on the floor was waiting for a signature.

(22b) The assistant forgot the letter was waiting for a signature.

The current experiments present clear evidence that the length of the ambiguous region, and in particular, the distance between the head of the ambiguous region and the disambiguating item, *does* affect reanalysis, suggesting that the Fodor and Inoue account requires modification to accommodate the present results.

A second critical question that must be addressed by models of reanalysis is the contrast between the easier reanalysis of NP/S structures as compared with the more difficult NP/Z reanalyses as demonstrated by the data of Sturt et al. (1999). (Recall that NP/S reanalyses are those where the ambiguous main verb can take either an NP or S complement, while NP/Z reanalyses are those where the main verb is ambiguous between a transitive structure with an NP complement or an intransitive structure with no (Zero) complement.) Consider again the NP/Z reanalysis in sentence (23).

(23) Before the woman visited the doctor had been drinking quite a lot.

According to Fodor and Inoue, this reanalysis is difficult because of the lack of a grammatical dependency between the disambiguating symptom *had* and the NP *the doctor*, as illustrated in Fig. 10. This forces the parser to resort to a technique they refer to as *theft*, in which adjacent items are stolen at the word level, without respecting existing constituent structure. This creates a tu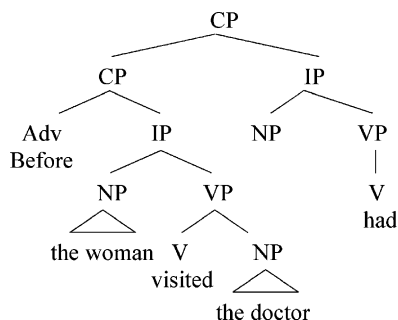g-of-war with *visited* and *had* both trying to assign their theta-roles to the same NP. Fodor and Inoue (1998) refer to this situation as an instance of the Thematic Overlay Effect, in which the parser resists any revision that reassigns items already associated with a certain thematic role unless it can substitute others for that role, or else eliminate it entirely by overlaying a new argument structure over the old one. These operations are under the purview of the *adjust* operator, which we have already seen does not apply in this sentence, so the tug-of-war cannot be resolved and reanalysis fails.

Thus, the Fodor and Inoue account must appeal to both an *adjust* and a *theft* operator to account for the difference between NP/S and NP/Z reanalyses. In contrast, a cue-based retrieval parser handles NP/Z reanalyses in the same way as it handles NP/S reanalyses, but unlike the NP/S case, the retrieval cues necessary to complete the NP/Z reanalysis are not available. Consider what happens in sentence (23). When the verb *had* occurs, two kinds of retrieval cues are set: retrieval cues for an item looking for a clause, and retrieval cues for a subject. The former cues are satisfied by the adjunct clause *Before the woman visited the doctor*, which expects to modify a main clause, and so *had* is integrated into the parse tree as in Fig. 10. The subject cues match both NPs *the doctor* and *the woman*, though neither are perfect candidates because they are already dependents in the parse tree. Assuming the higher activation of *the doctor* results in its retrieval, it will be attached as the subject of *had* (an "attach-anyway" operation (Fodor & Inoue, 1998)) resulting in the structure illustrated in Fig. 11. This structure is not a well-formed parse tree because *the doctor* is attached as both the subject of *had* and the object of *visited*. The only way this structure could be made acceptable is if the alternative non-transitive lexical frame of the verb *visited* were substituted for the transitive frame, but the incoming word *had* provides no retrieval cues for accessing it.

This accounts for the Sturt et al. (1999) evidence that these reanalyses are much more difficult than the NP/S reanalysis, in which the verb *has* provides all the necessary retrieval cues, and does not need to posit any additional mechanisms to do so. Thus, the retrieval theory provides a unified explanation of the head position

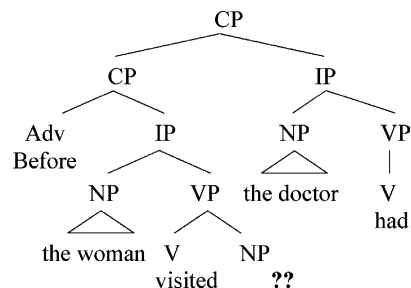Fig. 10. Attachment of disambiguating item into parse tree for NP/Z reanalyses.

Fig. 11. End state for NP/Z reanalyses according to the cue-based retrieval model.

effects in NP/S ambiguities, as well as the relative ease of NP/S ambiguities in contrast to NP/Z ambiguities. The explanation has much in common with the Fodor and Inoue model, in that it depends on the efficacy of the cues provided by the disambiguating signal. The key difference, however, is that that the cues are construed as retrieval cues, which are employed by the same retrieval mechanism thought to support initial structure building. Consequently, the explanation based on cue-based retrieval requires fewer mechanisms, and these mechanisms are independently motivated by the memory literature.

The cue-based retrieval approach to reanalysis can also account for recent data suggesting that reanalysis occurs "as a last resort" (Schneider & Phillips, 2001; Sturt et al., 2001). Using sentences similar to (24), both Sturt et al. and Schneider and Phillips report evidence that the verb *wrote* is attached high in the tree (with *woman* as its subject) despite strong bias from verbs like *knows* that it should head a complement clause and hence take *funny man* as its subject.

(24) The creative woman who knows the funny man wrote some comedy sketches [herself/himself] about the amusing escapades [she had seen/wants to publish them].

Schneider and Phillips explain these results by suggesting that the low attachment of *wrote* is not considered because the existing commitment of *the funny man* as the object of *knows* makes it unavailable to serve as the subject of the sentential complement structure. This is exactly the prediction made by the cue-based retrieval account. When the verb *knows* occurs, it sets retrieval cues for either an NP object or a sentential object. The occurrence of the *funny man* without an initial complementizer causes the NP object cues to be satisfied, and it is assigned object Case. Then, when *wrote* occurs, it can be incorporated into the tree in two ways- either as a matrix verb or as a subordinate verb. This produces two types of retrieval cues: those that initiate a search for a nominative marked NP and those that initiate a search for a verb that can host a subordinate clause. What it finds is the NP *woman who knows the funny man* since this is the only active structure available to match its retrieval cues, the alternative lexical frame of *knows* having decayed since the occurrence of *funny man*. Thus, the NP *woman* is retrieved as the dependent for *wrote* and a high attachment is made, which is consistent with the tendency of RALR to use available constituents and restructure only when no other possibilities are present. Cue-based reanalysis is triggered only when the occurrence of *himself* sets a retrieval cue for a masculine subject.

Finally, the cue-based retrieval parser can account for effects that have been assumed to be due to other factors. For example, Frazier and Clifton (1998) have introduced the notion of *visibility*, which they suggest is a property affecting both the parser and the reanalysis mechanism. Using the examples in (25a)–(25b), taken from Frazier and Clifton (1998), they explain the relative ease of (25a) over (25b) in terms of the increased depth of the right-branching material in the ambiguous region. They suggest that as the right-branch becomes deeper, many more recent nodes have been postulated and these cause the head of the phrase required for attachment of *and/had cost* to become *invisible*, and hence unavailable.

(25a) Unfortunately, [the space capsule everyone objected to] had been built by NASA *and cost* the taxpayers a fortune.
(25b) Unfortunately, [the space capsule everyone objected to after it was built by NASA] *had cost* the taxpayers a fortune.

As presently formulated, visibility does not accommodate the results observed in the present experiments. Our high interference and low interference sentences are matched for depth of the right-branch, but nevertheless the high interference sentences are judged to be ungrammatical significantly more often than the low interference sentences. We therefore concluded that recency (i.e., visibility) is not the cause of the contrast, but rather it is the interfering structure of the intervening region, together with the cue-based retrieval parser, which makes a required node inaccessible. Notably, interference effects can also be observed in the Frazier and Clifton sentences. Sentence (25b) would be considered a higher interference sentence on our account because the phrase enclosed in brackets includes at least three possible NPs which could fulfill the retrieval cues from *had cost* (i.e., *capsule*, *everyone*, *it*). In contrast, (25a) contains only two possible role fillers (i.e., *capsule*, *everyone*) for the subject role sought by *cost*, supplying less interference overall for completing the long distance attachment.

A number of other accounts have also been suggested for describing the processing difficulties in sentences such as (25a and 25b), as well as the contrast between the low and high interference sentences tested in the current experiments. These accounts have focused on a variety of explanatory factors, including difficulties caused by the need to maintain too many incomplete constituents (e.g., Abney & Johnson, 1991; Chomsky & Miller, 1963; Kimball, 1973; Yngve, 1960), too many self-embeddings (e.g., Gibson & Thomas, 1999; G. Miller & Isard, 1964; Miller & Chomsky, 1963), or too many incomplete dependencies (e.g., Gibson, 1998; Lewis, 1996). The main advantage of the cue-based retrieval account is that it does not require postulating additional principles or limits on the workings of the parser. Rather, our account unifies facts about language processing with evidence about the retrieval mechanism inherent in the working memory system, and highlights questions about the nature of the interaction between

the structure building property of human language parser and the retrieval mechanism used in other working memory tasks, such as the list learning tasks in which interference effects were first observed.

In addition, this account presents a challenge for the alternative accounts of complexity effects mentioned above, because they must also explain other types of interference effects observed in embedded sentences. For example, Van Dyke (2002, 2003) observed semantic interference effects in sentences such as (26a) and (27a) as compared with the non semantically interfering sentences in (26b) and (27b).

(26a) The executive assistant forgot that the client who had arrived for the important visitor departed.
(26b) The executive assistant forgot that the client who had arrived with the important meeting departed.
(27a) The executive assistant forgot that the client who implied that the visitor was important departed.
(27b) The executive assistant forgot that the client who implied that the meeting was important departed.

In (26 and 27a) the retrieval cues provided by the verb *departed* are not sufficient to identify *client* as the correct subject of the embedded clause because there is a more recent NP which also fits the semantic cues provided by *departed*. Alternative accounts which appeal to costs associated with clause closure or storage limitations do not accommodate these results, since the number of clauses and storage requirements are identical in both (a) and (b) versions of these sentences.

In light of these and other cross-linguistic data demonstrating a range of interference effects due to factors such as category membership (Gordon et al., 2001), positional similarity (Lewis & Nakayama, 2002; Uehara & Bradley, 1996), and phonological similarity (Lee & Nakayama, 2003), we are encouraged that the approach presented here will provide a framework for exploring how the phenomenon of similarity-based interference illuminates the memory architecture and associated processing structures necessary for human sentence comprehension.

Conclusion

In this paper, we have examined a property of *structure*, i.e., interference, which we have shown affects both parsing and reanalysis via a *single* mechanism, i.e., cue-based retrieval, which is responsible for both identifying grammatical dependencies and initiating structural change. This mechanism is sensitive to structural ambiguities via the retrieval cues it must use to identify items needed to fill grammatical dependencies. When these cues do not permit unambiguously identifying appropriate

attachment sites and/or fillers, then processing becomes more difficult, as was the case with the high interference structures examined here. Critically, this difficulty appears in both unambiguous and ambiguous sentences, suggesting that the cue-based retrieval mechanism operates independently of those processes involved during the revision itself. Revision processes, on the other hand, are influenced by a separate factor: decay, which becomes critical when reanalysis requires the retrieval of lexical frames which have been shifted out of the focus of attention due to distributional facts in the language. We have shown that this effect supports the previous work of Ferreira and Henderson (1991b), but have extended this work by illustrating how this effect arises via cue-based retrievals, which occur during the repair process and not during the processes involved in creating the globally correct attachment. Consequently, the mechanism of cue-based retrieval arises as a critical component of structure-building in both ambiguous and unambiguous contexts, and provides an important framework for investigating the contribution of the memory mechanisms that support language processing more generally.

Appendix

Materials for Experiments 1–4. Experiment 1 used only ambiguous versions of these sentences, and did not include a *that* in the embedded clause. Experiment 2 used both ambiguous and unambiguous versions of these sentences, but without a *that* in the embedded clause. Experiments 3 and 4 used the materials as presented here. Condition manipulations are shown in brackets. Sentences with a short intervening region contained only the NP (first in brackets). The high interference conditions contained the second clause in the brackets; the low interference conditions contained the third clause in the brackets.

1. The executive assistant forgot (that) [the letter\the letter which revealed that the mayor was responsible\the letter which had fallen on the floor] was waiting for a signature.
2. The frightened boy understood (that) [the man\the man who said that the townspeople were dangerous\the man who was swimming near the dock] was paranoid about dying.

3. The school principal remembered (that) [the teacher\the teacher who guesses that the boys were cheating\the teacher who was designing the new curriculum] would come to his office.
4. The television anchorman reported (that) [the story\the story which means that the war is over\the story which had exposed the government conspiracy] had broken this morning.
5. The news editor wrote (that) [the article\the article which suggested that the actress was shy\the article which was rejected from the paper] was a joke to the public.
6. The commanding officers recognized (that) [the troops\the troops who assumed that the snipers were captured\the troops who were drilling at the academy] would win the battle.
7. The greedy dictator denied (that) [the law\the law which said that the revolution was legal\the law which would reduce his absolute power] was justified by the crisis.
8. The math professor proved (that) [the theorem\the theorem which indicates that the algorithm would work\the theorem which was printed in the textbook] was easy for the students.
9. The hotdog vendor noticed (that) [the boy\the boy who thought that the fries were burnt\the boy who was annoying the other customers] was crying on the curb.
10. The rock band recognized (that) [the manager\the manager who said that the room was ready\the manager who congratulated them about their record] could negotiate a deal.
11. The drama critic wrote (that) [the play\the play which demonstrated that elaborate scenery is unnecessary\the play which starred the talented young actors] would succeed on Broadway.
12. The army scout noticed (that) [the settlers\the settlers who realized that the settlement was risky\the settlers who were starting a new town] were afraid of bears.
13. The efficient conductor knows (that) [the man\the man who thinks that the train is cold\the man who is traveling with the boombox] is causing trouble in the lounge.
14. The drama student envisioned (that) [the actor\the actor who said that the play was challenging\the actor who is visiting the acting class] would speak for an hour.
15. The horse trainer felt (that) [the horse\the horse who showed that the record was beatable\the horse who led in the first lap] was a winner for sure.
16. The police officer reported (that) [the crime\the crime which indicated that bank security was weak\the crime which occurred at the amusement park] was solved last week.
17. The senior senator proclaimed (that) [the rights\the rights which prove that the nation is free\the rights which protect us from illegal search] should be guaranteed to all.
18. The business manager found (that) [the company\the company that recognizes that the competition is tough\the company which created the bright colorful ads] was very creative on the Web.
19. The science teacher heard (that) [the student\the student who said that the assignment was difficult\the student who usually gets the highest score] would be late this afternoon.
20. The insurance inspector forgot (that) [the letter\the letter which said that the company was bankrupt\the letter which was addressed to the manager] was so full of errors.
21. The crew manager saw (that) [the handyman\the handyman who thought that the project was finished\the handyman who bought the hammer and nails] was smoking near the office.
22. The intelligent juror recalled (that) [the witness\the witness who assumed that the plaintiff was innocent\the witness who was watching from the back] would cause trouble in the future.
23. The young mother remembered (that) [the dog\the dog which thinks that the chain is short\the dog which was running in the backyard] would be dangerous for the baby.
24. The university president proclaimed (that) [the news\the news which said that classes would be cancelled\the news which got passed around the dorms] was based on inaccurate information.
25. The store clerk found (that) [the customer\the customer who said that the size was wrong\the customer who was buying expensive silver shoes] was waiting at the counter.
26. The bus driver saw (that) [the lady\the lady who complained that the seats were uncomfortable\the lady who was sitting in the front] was asleep before the crash.
27. The interior decorator envisioned (that) [the rug\the rug which shows that the executive is wealthy\the rug which was hanging in the store] would complete the room.
28. The district attorney understood (that) [the case\the case which claims that the plaintiff was injured\the case which accused the angry old man] was weak but defensible.
29. The political scientist read (that) [the book\the book which revealed that the embargo had failed\the book which was published about the revolution] was banned in Russia.
30. The baseball coach knows (that) [the player\the player who thinks that the team is losing\the player who practices for three extra hours] is good in left field.
31. The concerned doctor recalled (that) [the specialist\the specialist who thinks that the disease is curable\the specialist who is known for successful treatments] would be good for the patient.
32. The young lawyer proved (that) [the case\the case which claimed that the company was innocent\the case which was presented against the plaintiff] was unfair to the client.
33. My older sister heard (that) [the secretary\the secretary who thinks that the rules are effective\the secretary who is guiding the tour group] is doing too much work.
34. The faculty advisor denied (that) [the complaint\the complaint which proposed that the strikers were guilty\the complaint which attacked the new student organization] was justified at all.
35. The English major read (that) [the play\the play that showed that the hero was vulnerable\the play which featured women amateurs on stage] was a classic of Irish literature.
36. The experienced doctor felt (that) [the child\the child who thought that the illness was bronchitis\the child who complained about sharp stomach pains] was allergic to milk.

References

Abney, S. P., & Johnson, M. (1991). Memory requirements and local ambiguities of parsing strategies. *Journal of Psycholinguistic Research, 20*, 233–250.

Altmann, E. M., & Gray, W. D. (2002). Forgetting to remember: The functional relationship of decay and interference. *Psychological Science, 13*(1), 27–33.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Anderson, J. R., & Matessa, M. (1997). A production system theory of serial memory. *Psychological Review, 104*(4), 728–748.

Bailey, K. G. D., & Ferreira, F. (in press). Disfluencies affect the parsing of garden-path sentences. *Journal of Memory and Language*.

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.

Chomsky, N., & Miller, G. A. (1963). Introduction to the formal analysis of natural languages. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2, pp. 269–321). New York: Wiley.

Christianson, K., Hollingworth, A., Halliwell, J., & Ferreira, F. (2001). Thematic roles assigned along the garden-path linger. *Cognitive Psychology, 42*, 368–407.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review, 3*(1), 37–60.

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

Earley, J. C. (1970). An efficient context-free parsing algorithm. *Communications of the ACM, 13*(2), 94–102.

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

Ferreira, F., & Henderson, J. M. (1991a). How is verb information used during syntactic parsing? In G. B. Simpson (Ed.), *Understanding word and sentence* (pp. 305–330). Amsterdam: Elsevier, North-Holland.

Ferreira, F., & Henderson, J. M. (1991b). Recovery from misanalyses of garden path sentences. *Journal of Memory and Language, 30*, 725–745.

Ferreira, F., & Henderson, J. M. (1993). Reading processes during syntactic analysis and reanalysis. *Canadian Journal of Experimental Psychology, 47*(2), 247–275.

Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry, 9*, 427–473.

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

Fodor, J. D., & Inoue, A. (1994). The diagnosis and cure of garden paths. *Journal of Psycholinguistic Research, 23*(4), 405–432.

Fodor, J. D., & Inoue, A. (1998). Attach Anyway. In J. D. Fodor, & F. Ferreira (Eds.), *Reanalysis in Sentence Processing*. Dordrecht: Kluwer.

Fodor, J. D., & Inoue, A. (2000). Garden path re-analysis: Attach (anyway) and revision as last resort? In M. D. Vincenzi, & V. Lombardo (Eds.), *Cross-linguistic perspective on language processing*. Dordrecht: Kluwer.

Ford, M., Bresnan, J., & Kaplan, R. (1982). A competence-based theory of syntactic closure. In J. Bresnan (Ed.), *The mental representation of grammatical relations*. Cambridge: MIT Press.

Francis, W. N., & Kucera, J. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.

Frazier, L. (1978). On comprehending sentences: Syntactic parsing strategies. Unpublished doctoral dissertation, University of Connecticut, Storrs.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance, Vol. XII: The psychology of reading* (pp. 601–681). Hillsdale, NJ: Erlbaum.

Frazier, L., & Clifton, C. (1998). Sentence Reanalysis and Visibility. In J. D. Fodor, & F. Ferreira (Eds.), *Reanalysis in sentence processing*. Dordrecht: Kluwer.

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

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*(1), 58–93.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.

Gernsbacher, M. A., & Faust, M. (1990). The role of suppression in sentence comprehension. In G. B. Simpson (Ed.), *Understanding word and sentence* (pp. 97–128). Amsterdam: North-Holland.

Gibson, E. (1991). A computational theory of human linguistic processing: Memory limitations and processing breakdown. (Unpublished doctoral dissertation No. Center for Machine Translation Technical Report CMU-CMT-91-125.). Pittsburgh, PA: Carnegie Mellon University.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition, 68*, 1–76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz (Ed.), *Image, language, brain: Papers from the first mind articulation project symposium* (pp. 94–126). Cambridge, MA: MIT Press.

Gibson, E., & Pearlmutter, N. (2000). Distinguishing serial and parallel parsing. *Journal of Psycholinguistic Research, 29*(2), 231–240.

Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes, 14*(3), 225–248.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*, 1–65.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(6), 1411–1423.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science, 13*(5), 425–430.

Gorrell, P. (1995). *Syntax and parsing*. Cambridge, UK: Cambridge University Press.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers, 16*(96–101).

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95*(4), 528–551.

Hogaboam, T. W., & Perfetti, C. A. (1975). Lexical ambiguity and sentence comprehension. *Journal of Verbal Learning and Verbal Behavior, 14*, 265–274.

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

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

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*, 580–602.

Kintsch, W., & Mross, F. (1985). Context effects in word identification. *Journal of Memory and Language, 24*, 336–349.

Lee, S. H., & Nakayama, M. (2003, March 27–29). Effects of syntactic and phonological similarity in Korean center-embedding constructions. Paper presented at the 2003 CUNY Conference on Human Sentence Processing, Cambridge, MA.

Lewis, R. L. (1993). An architecturally-based theory of human sentence comprehension. Unpublished Doctoral dissertation, available as Technical Report CMU-CS-93-226 at reports@cs.cmu.edu, or Computer Science Department, CMU, 5000 Forbes Ave., Pittsburgh, PA 15213, Carnegie Mellon University, Pittsburgh, PA.

Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research, 25*(1), 93–115.

Lewis, R. L. (1998). Reanalysis and limited repair parsing: Leaping off the garden path. In J. D. Fodor, & F. Ferreira (Eds.), *Reanalysis in sentence processing*. Dordrecht: Kluwer.

Lewis, R. L. (2000). Falsifying serial and parallel parsing models: Empirical conundrums and an overlooked paradigm. *Journal of Psycholinguistic Research, 29*(2), 241–248.

Lewis, R. L., & Nakayama, M. (2002). Syntactic similarity effects of embeddings in Japanese. In M. Nakayama (Ed.), *Sentence processing in East Asian languages* (pp. 85–110). Stanford: CSLI.

Lewis, R. L., Van Dyke, J. A., Vasishth, S., & Nakayama, M. (submitted). Parsing as working memory retrieval. *Journal of Memory and Language*.

Lombardo, V. (1998). A computational model of recovery. In J. D. Fodor, & F. Ferreira (Eds.), *Reanalysis in sentence processing*. Dordrecht: Kluwer.

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

Marcus, M. P., Hindle, D., & Fleck, M. (1983). D-theory: Talking about talking about trees. Paper presented at the Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics, Cambridge, MA.

Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics, 19*(2), 313–330.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory. *Journal of Psycholinguistic Research, 29*(2), 111–123.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language, 48*(1), 67–91.

Meng, M., & Bader, M. (2000). Mode of disambiguation and garden-path strength: An investigation of subject-object ambiguities in German. *Language and Speech, 43*(1), 43–74.

Miller, G., & Isard, S. (1964). Free recall of self-embedded English sentences. *Information and Control, 7*, 292–303.

Miller, G. A., & Chomsky, N. (1963). Finitary models of language users. In D. R. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (vol. II, New York: Wiley.

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*(6), 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.

Neill, W. T., Hilliard, D. V., & Cooper, E. (1988). The detection of lexical ambiguity: Evidence for context-sensitive parallel access. *Journal of Memory and Language, 27*, 279–287.

Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition, 15*, 225–236.

Perfetti, C. A., & Tan, L. H. (1998). The time course of graphic, phonological, and semantic activation in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(1), 101–118.

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*(4), 940–961.

Rayner, K. (1977). Visual attention in reading: Eye movements reflect cognitive processes. *Memory & Cognition, 5*(4), 443–448.

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

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., 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., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 1188–1200.

Schneider, D., & Phillips, C. (2001). Grammatical search and reanalysis. *Journal of Memory and Language, 45*, 308–336.

Schneider, W. (1995). *Mel Professional (Version 2.0)*. Pittsburgh, PA: Psychology Software Tools, Inc..

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. L., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology, 14*, 489–537.

Stevenson, S. (1998). Parsing as incremental restructuring. In J. D. Fodor, & F. Ferreira (Eds.), *Reanalysis in sentence processing*. Dordrecht: Kluwer.

Sturt, P., & Crocker, M. (1998). Generalized monotonicity for reanalysis models. In J. D. Fodor, & F. Ferreira (Eds.), *Reanalysis in sentence processing*. Dordrecht: Kluwer.

Sturt, P., & Crocker, M. W. (1996). Monotonic syntactic processing: A cross-linguistic study of attachment and reanalysis. *Language and Cognitive Processes, 11*(5), 449–494.

Sturt, P., & Crocker, M. W. (1997). Thematic monotonicity. *Journal of Psycholinguistic Research, 26*(3), 297–322.

Sturt, P., Pickering, M. J., & Crocker, M. W. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language, 40*, 136–150.

Sturt, P., Pickering, M. J., Scheepers, C., & Crocker, M. W. (2001). The preservation of structure in language comprehension: Is reanalysis the last resort. *Journal of Memory and Language, 45*, 283–307.

Tabor, W., & Hutchins, S. (2003). Evidence for self-organized parsing: Digging in effects. Paper presented at the 2003 CUNY Conference on Human Sentence Processing, Cambridge, MA.

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

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints on sentence processing: Separating effects of lexical preference from garden paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(3), 528–553.

Uehara, K., & Bradley, D. (1996). The effect of -ga sequences on processing Japanese multiply center-embedded sentences. In B. Park, & J. Kim (Eds.), *Language, information and computation*. Seoul: Kyung Hee University.

Van Dyke, J. A. (2002). Retrieval effects in sentence parsing and interpretation. Unpublished Ph.D. Dissertation, University of Pittsburgh, Pittsburgh.

Van Dyke, J. A. (2003). Complexity effects are retrieval effects. Paper presented at the 16th Annual CUNY Conference on Human Sentence Processing, Cambridge, MA.

van Gompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language, 45*, 225–258.

Vasishth, S. (2002). Working memory in sentence comprehension: Processing Hindi center embeddings. Unpublished Ph.D. Dissertation, Ohio State University, Columbus, OH.

Warner, J., & Glass, A. L. (1987). Context and distance-to-disambiguation effects in ambiguity resolution: Evidence from grammaticality judgment of garden path sentences. *Journal of Memory and Language, 26*, 714–738.

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition, 85*(1), 79–112.

Weinberg, A. (1993). Parameters in the theory of sentence processing: Minimal commitment theory goes East. *Journal of Psycholinguistic Research, 22*(3), 339–364.

Yngve, V. H. (1960). A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society, 104*, 444–466.