RETRIEVAL EFFECTS IN SENTENCE PARSING AND INTERPRETATION

by

Julie A. Van Dyke

Bachelor of Arts, University of Delaware, 1991

Master of Science, Carnegie Mellon University, 1996

Submitted to the Graduate Faculty of

Arts & Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2002

UNIVERSITY OF PITTSBURGH

FACULTY OF ARTS AND SCIENCES

This dissertation was presented

by

Julie A. Van Dyke

It was defended on

August 7, 2002

and approved by

Julie A. Fiez

Alan Juffs

Erik D. Reichle

James F. Voss

Charles A. Perfetti
Dissertation Director

Charles A. Perfetti, Advisor

RETRIEVAL EFFECTS IN SENTENCE PARSING AND INTERPRETATION

Julie A. Van Dyke, PhD

University of Pittsburgh, 2002

This dissertation is an investigation into the memory mechanisms that support parsing and how they constrain parsing success. Constraints can occur from two directions: the ability to *store* items participating in grammatical dependencies, or else the dependencies themselves, and the ability to *retrieve* items required for creating grammatical dependencies. A review of available evidence regarding memory constraints on parsing suggests that although attention has more often been focused on the *storage* constraints on parsing, the evidence is actually more consistent with a *retrieval* account of parsing breakdown. I present a framework for investigating retrieval effects in parsing, based on existing models of associative memory retrieval and illustrate how the work traditionally carried out by the parser can be understood as an instance of working memory retrieval. Five experiments present new evidence in support of the retrieval approach. Two of these illustrate effects of semantic interference, in which semantic properties of candidate NPs affect the ability to complete grammatical dependencies, and the remaining three illustrate effects of referential interference, in which the availability of items in the situation model affect parsing success. It is argued that storage accounts of parsing breakdown cannot account for these results, promoting the conclusion that an associative memory mechanism underlies both parsing and sentence interpretation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Identifying Retrieval Effects

1.1 Introduction

Sentence comprehension requires the ability to identify grammatical dependencies among words in a sentence. Traditionally, the mechanism thought to be responsible for this is the *parser*, which is typically thought to construct syntactic structure via the application of grammatical knowledge. Yet despite access to this grammatical knowledge, empirical observation has shown that certain types of sentences are not at all easy for the parser to process, with some leading to complete comprehension breakdown, even when the grammar of the language predicts them to be comprehensible.

This dissertation is an investigation into the memory mechanisms that support parsing and how they constrain parsing success. From this point of view, parsing can be constrained from two directions: the ability to *store* items participating in grammatical dependencies, or else the dependencies themselves, and the ability to *retrieve* items required for creating grammatical dependencies. In this chapter, I will review the available evidence regarding memory constraints on parsing, with special attention to where in the sentence these effects occur, and what are the concurrent demands on memory. Based on this review, I will conclude that although attention has more often been focused on the *storage* constraints on parsing, the evidence is actually more consistent with a *retrieval* account of parsing breakdown. I will then present a framework for investigating retrieval effects in parsing, based on existing models of associative memory retrieval. I will illustrate how the work traditionally carried out by the parser can be understood as an instance of working memory retrieval, and highlight three predictions these models make about when parsing breakdown will occur. The remainder of the dissertation will present empirical tests of these predictions, attesting to the suitability of the retrieval account and highlighting implications of this account for an understanding of the role of the parser in sentence interpretation in general.

1.2 Role of Working Memory in Parsing

The fact that language is experienced incrementally promotes a focus on the *storage* aspect of memory constraints because grammatically dependent items can often be separated by several constituents. According to the storage view (cf. Carpenter, Miyake, and Just (1994) for a review), comprehension breakdown occurs when capacity limitations are exceeded. The classic comparison for illustrating this effect is the contrast between subject relative clauses such as Example 1.1 as compared with object relative clauses such as Example 1.2.

Example 1.1: The reporter that attacked the senator admitted the error.

Example 1.2: The reporter that the senator attacked admitted the error. (King & Just, 1991)

The object relative in Example 1.2 is argued to be more complex than Example 1.1 because it putatively requires more working memory resources to process, a conclusion suggested by reliable differences produced in a wide range of experimental paradigms including reading times (Ford, 1983; King & Just, 1991), eye-movement studies (Holmes & O'Regan, 1981), response-accuracy to comprehension questions (Wanner & Maratsos, 1978), ERPs (Vos, Gunter, Schriefers, & Friederici, 2001), and measures of blood flow volume (Just, Carpenter, Keller, Eddy, & Thulborn, 1996b; Stromswold, Caplan, Alpert, & Rauch, 1996) (cf. Gibson (1998) for a review of related literature.) As noted by Gibson (1998), there is little consensus regarding exactly why the object-relative structures are more difficult, but all of the available proposals have centered on explaining the effect in terms of a *processing overload*. Some measure the overload in terms of too many incomplete constituents (e.g. (Abney & Johnson, 1991; Chomsky & Miller, 1963; Kimball, 1973; Yngve, 1960)), too many embeddings (e.g., (Gibson & Thomas, 1999; Miller & Isard, 1964; Miller & Chomsky, 1963)), too many perspective shifts (e.g., (Bever, 1970; MacWhinney & Pleh, 1988)), or too many incomplete dependencies (e.g., (Gibson, 1998; Lewis, 1996)).

1.2.1 Just and Carpenter's Capacity Theory of Comprehension

In light of research on the affects of working memory load and individual memory span on comprehension, this focus on the limiting effects of working memory capacity seems correct. The logic underlying such research is as follows: if comprehension is limited by working memory capacity, then other factors which limit working memory, such as low working memory span or high working memory load, should affect comprehension even further. In a highly influential paper King and Just (1991) tested this hypothesis directly, by comparing high and low working memory span subjects' comprehension of the sentences in Example 1.1 & Example 1.2 in a variable memory load task. Subjects' memory span was assessed using the Daneman and Carpenter (1980) reading span test, which requires subjects to read an increasingly large set of sentences while remembering the last word of each sentence in the set. For example, subjects might read three sentences, and at the end of the third sentence be asked to recall the last word of that sentence as well as the last word of each of the preceding two sentences. This would be a set size of three (i.e. three sentences to read and three last words to remember), and the subject would be given five trials in this set size, before moving up to a set size of four (i.e. four sentences to read and four last words to remember). Reading span is defined as the largest set for which the subject has perfect recall for three of the five trials in a given set size.

Having determined which subjects are high and low span subjects, King and Just then asked both groups to read subject and object relative clauses in a moving window paradigm. These sentences were shown as the final sentence in sets of 1, 2, or 3 sentences, and subjects were asked to remember the final words of each of the preceding sentences (note that in one-sentence sets, the critical relative clause sentence was the only sentence shown). At the conclusion of reading the critical relative clause sentence, subjects were asked to recall the words they were remembering, and then to answer a comprehension question about the critical relative clause sentence. As expected, they found that overall recall was worse for low span subjects than for high span subjects, but also that the size of the memory load differentially affected the low span subjects, making recall in the higher load 3-sentence condition more difficult than in the 1-sentence conditions. They interpreted this result as evidence that higher memory span subjects had enough capacity to manage the working memory load without it interfering with comprehension, whereas the lower span subjects who started out with fewer memory resources did not have the residual memory capacity required for comprehension when the memory load was high. This effect of memory load also interacted with the type of relative clause in their experiment. Recall after the object relative sentence was much worse in the high load conditions than if the final sentence was a subject relative, supporting the hypothesis that object relative clauses are difficult because they tax working memory capacity in a way that subject relative clauses do not.

The relationship between memory load and comprehension can be seen more directly in the King and Just comprehension question data. For 1-sentence trials, they found an interaction between relative clause type and memory capacity, such that the low span subjects had much more difficulty (22% more) with the object relative clauses as compared to the high span subjects. In contrast, the difference between the high and low span subjects on subject relative clauses was only 7%. According the King and Just account, this is because of the inherent ease of processing the less complex subject relative, while memory span was more critical for understanding the object relative sentences.

These findings have led Just and Carpenter and colleagues to propose the capacity theory of comprehension (Daneman & Carpenter, 1980; Just & Carpenter, 1992; King & Just, 1991; MacDonald, Just, & Carpenter, 1992), which proposes a unitary working memory capacity for all verbal processing, much in the tradition of Baddeley's componential view of working memory (Baddeley, 1992). The Just and Carpenter view places verbal working memory within the domain of Baddeley's central executive, which allocates resources for the job of language (and all verbal) processing (Just & Carpenter, 1992). As such, a single pool of memory resources is responsible for the verbal processing required by a task such as the Daneman and Carpenter (1980) working memory span task as well as building the syntactic structure necessary for comprehension. Such a memory model leads to the predictions we have already seen: that reading span and memory load can affect language comprehension processes such that subjects who perform poorly on the memory span task will also be disadvantaged at comprehending more complex syntactic structures (e.g., object relative clauses). A second type of complexity that this model suggests will reveal individual differences stems not just from the syntactic structure of the sentence, but from the processing complexity of the sentence as a whole. As we have seen above, this model predicts that an increased memory load depletes the available pool of memory resources, making structure building more difficult. One way of testing this hypothesis is to manipulate memory load using an *extrinsic* memory task, such as the King and Just (1991) procedure we have already seen, which requires subjects to remember and recall a series of words while they are processing the target sentences. A second test of the capacity hypothesis involves the use of sentences with temporary ambiguities, which provides an *intrinsic* manipulation of memory load because some amount of additional processing is required to resolve the ambiguity. For example, the sentences in Example 1.3 and Example 1.4 are ambiguous in the region of *soldiers warned about the dangers* between a reduced relative clause structure and a main verb reading of *warned*.

Example 1.3: The experienced soldiers warned about the dangers before the midnight raid.

Example 1.4: The experienced soldiers warned about the dangers conducted the midnight raid.
(MacDonald et al., 1992)

MacDonald et al. (1992) have suggested that whether the reader maintains both possible interpretations depends critically on whether that reader has enough working memory resources available to do so. Thus, a high span reader will maintain both interpretations because s/he has the resources to do so, resulting in high comprehension rates regardless of whether the sentence is resolved to its main verb or reduced relative reading. In contrast, a low span reader will have insufficient resources to maintain both interpretations, and so will have good comprehension when the ambiguity is resolved to its more preferred main verb reading, but not when the resolution is to the reduced relative reading. While this prediction is interesting, evidence of this sort is not sufficient to support the capacity theory of language comprehension because better comprehension by high span readers may be due to better reading strategies that are unrelated to the allocation of memory.

According to MacDonald et al., the data that distinguishes an account like this from theirs is whether high span readers will have *more* difficulty with the ambiguous sentences than low span readers. This is because the costs of maintaining more than one interpretation would cause high span readers to be unable to allocate processing resources for other types of comprehension operations, such as higher level reference resolution or elaboration, which they would have to catch up on once the ambiguity was resolved. Thus, high span readers should show increased reading times for ambiguous sentences despite high comprehension rates, while low span readers should show low comprehension rates but higher reading times compared to the high span readers. This is precisely the pattern that MacDonald et al. found in their data, when comparing the ambiguous sentences in Example 1.3 and Example 1.4 to unambiguous controls shown in Example 1.5 and Example 1.6.

Example 1.5: The experienced soldiers spoke about the dangers before the midnight raid.

Example 1.6: The experienced soldiers who were told about the dangers conducted the midnight raid. (MacDonald et al., 1992)

Their results are unconvincing, however, because the increased reading times for the high span readers occurred at the final word of the sentence, even for the reduced relative conditions where the disambiguation occurred several words prior (e.g. the verb *conducted* in Example 1.4). While it is true that for the main verb conditions such as those in Example 1.3 the disambiguation occurs at the period in the sentence, attributing this affect to costs associated with maintaining an increased memory load is problematic because the end of the sentence is known to carry its own wrap-up costs (cf. (King & Just, 1991) and (Just & Carpenter, 1980) who do not analyze end-of-sentence data for exactly this reason). Since high span readers may simply be better at end-of-sentence operations because they are better readers overall, these results do not provide compelling evidence that high span readers are maintaining both the preferred and unpreferred interpretations of the ambiguous main verb sentences, and therefore that any processing differences are due to memory capacity limitations. Similarly, if high span readers paid a cost for maintaining multiple representations which appeared as "catch-up" work after the ambiguity was resolved, then this cost should occur at the point of the resolution in the relative clause sentences, and not at the final word. In fact the high span readers had *faster* and not slower reading times at the point of disambiguation in these sentences, data which again calls into question the capacity constraint account of individual differences in processing these temporary ambiguities.

1.2.2 Waters and Caplan's Language Specific Working Memory model

There are a number of difficulties with the application of the Just and Carpenter Capacity model to the processing of temporary ambiguities, including the fundamental assumption that both interpretations are automatically generated, even when memory resources are plentiful (cf. (Fodor & Inoue, 2000; Frazier, 1987; Frazier & Clifton, 1996)). Indeed, in an attempt to replicate the MacDonald et al. finding, Waters and Caplan (1996b) did not find the crucial interaction between memory load and sentence type, even in rapidly presented paradigms (RSVP) that increased the cognitive load during comprehension. This failure to replicate the MacDonald et al. (1992) finding is one of several points on which Waters and Caplan (1996a; 1996b) have criticized Just and Carpenter's Capacity theory. Their central claim is that the automatic processes required for sentence comprehension, including lexical access, the construction of syntactic structure, the assignment of thematic roles, and determination of discourse-level semantic values draw from a distinct language-specific resource pool which is not identical to the pool required for verbally mediated tasks which are conscious and controlled, such as the Daneman and Carpenter memory span task. In this way, they have decoupled the capacity for language processing from that measured by general verbal working memory tasks, and rejected the prediction that lies at the core of the Just and Carpenter theory: that low working memory capacity will affect language processing (Caplan & Waters, 1999). The primary support for this hypothesis comes from studies with patient populations which show reduced working memory capacity, but intact language processing of complex sentences. For example, Waters and Caplan (1991) tested a patient B.O. who had a memory span of only two or three items on recall and probe-recognition tasks, and one item when measured using the Daneman and Carpenter task, however, she performed as well as normals on syntactically complex relative clauses and garden path sentences. In addition, Rochon, Waters, and Caplan (1994) and Waters, Rochon, and Caplan (1995) found that Alzheimer's patients who had a Daneman and Carpenter memory span of one or less, were not poorer than controls at comprehending complex sentences. A similar result was obtained in a study with Parkinson's disease patients who showed reduced working memory spans as compared to controls, but no differential impairment on object relative sentences versus subject relatives as compared to controls (Caplan & Waters, 1999). These data, taken together with weaknesses in the data supporting the Just and Carpenter Capacity Theory (cf. Waters and Caplan (1996a)), seem to support the view that a distinct language-specific memory resource supports language comprehension, and that the operations which draw on this language component appear to operate automatically, so that even in the face of reduced overall working memory span, language comprehension remains intact.

It is precisely this appeal to automaticity that Just, Carpenter, and Keller (1996a) have suggested makes the Waters and Caplan proposal of a language-specific memory resource untenable. They claimed that the automaticity of a process can vary depending on its context, and so the fact that certain processes appear unaffected by decreased memory load does not lead to the conclusion that those processes do not draw on a single working memory resource. For example, Just et al. (1996a) suggested that lexical access for a short, familiar word may be fast and automatic, but when a word is very infrequent, lexical access can take much longer than in normal reading (cf. (Just & Carpenter, 1980)). They point out that the Waters and Caplan proposal has difficulty accounting for the latter case because lexical access is supposed to be within the scope of their automatic, language-specific processing resource. However, the fact that lexical access (or any other process) may be slow or unsuccessful does not imply that it occurs with conscious problem solving of the sort that draws on the same memory resource as would the Daneman and Carpenter task. Rather, it may simply be a function of whether the process' completion criteria are met, which in the case of lexical access, is a function of how easily a word can be retrieved. Consequently, this criticism does not present a convincing argument against a language-specific resource, but merely reinforces the observation that focusing solely on the storage aspect of working memory yields an inadequate explanation of processing difficulty.

1.2.3 Further Specialization of Working Memory: Martin's View

The chief distinction between the Waters and Caplan and the Just and Carpenter views is the degree to which working memory is fractionated into a specific component supporting language processing. At the theoretical level, the drive towards fractionation reflects the influence of the modularity hypothesis (Fodor, 1983), which proposes a set of informationally encapsulated processors which accepts only a limited range of inputs and operates automatically on that input to produce a specific type of output. We have seen that in the proposal of Waters and Caplan, this specialization includes all those processors involved in transforming the acoustic signal into a discourse-coherent semantic representation, including acoustic-phonetic conversion, lexical access, recognition of intonational contours, syntactic processing, and determination of discourse-level semantic values such as topic, focus, coreference, causality, and temporal order (Caplan & Waters, 1999). They have argued that all this processing occurs using the same representational language, and is highly over-practiced, giving it the automatic character required of an encapsulated module. However, the degree to which this type of processing is actually informationally encapsulated has been debated widely in the literature, largely on the basis of the extent to which semantic and pragmatic information influences syntactic processing (e.g., (Britt, Perfetti, Garrod, & Rayner, 1992; Ferreira & Clifton, 1986; MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell, Tanenhaus, & Garnsey, 1994)). The typical approach in these studies is to manipulate the amount of pragmatic facilitation available for the interpretation of ambiguous sentences, in an attempt to forestall what would otherwise be a garden path. For example, Ferreira and Clifton (1986) tested the temporarily ambiguous sentences such as those in Example 1.7 and Example 1.8 as compared with the unambiguous sentences in Example 1.9 and Example 1.10. Note that prior to the *by*-phrase in Example 1.8, this sentence could have either a main clause reading of *examined*, with *defendant* as the subject, or a relative clause reading such as that in Example 1.10. However, this is not the case for Example 1.7 since *evidence* is not an appropriate subject for *examined* as a main verb.

Example 1.7: The evidence examined by the lawyer shocked the jury.

Example 1.8: The defendant examined by the lawyer shocked the jury.

Example 1.9: The evidence that was examined by the lawyer shocked the jury.

Example 1.10: The defendant that was examined by the lawyer shocked the jury.

They found that subjects' fixation durations were longer on the *by the lawyer* phrase in both Example 1.7 and Example 1.8 suggesting that subjects do not use the available animacy cues to inform syntactic processing, perhaps due to architectural restrictions on the type of information available to the syntactic processor. Follow-up research has shown that the extent to which semantic information is used in structure building depends on the strength of the semantic constraint (Trueswell et al., 1994). When the constraint is strong, as in Example 1.7, then the semantic effect *is* in fact found; however, when the constraint is weaker, as was the case with a high proportion of the items in the Ferreira and Clifton study, then the semantic effect is not found.

The Waters and Caplan proposal appears to draw the line for encapsulation between the syntactic processing required for structure-building and the semantic/pragmatic processing required for interpreting lexical semantic properties of words that would determine their suitability as role fillers in experiments such as these (Caplan & Waters, 1999; Waters & Caplan, 1996a). In light of the data presented in the Trueswell et al. (1994) study, Waters and Caplan have suggested that pure encapsulation is not required in order for a set of processes to be cognitively separable or to rely on a specialized resource pool (Caplan, 1985). While this may appear to weaken their position theoretically, data of Martin and colleagues from brain-damaged patients with a disruption of lexical-semantic processing, but not that of syntactic or phonological processing seen to support the distinction they are attempting to preserve (Martin & Feher, 1990; Martin & Romani, 1994; Martin, Shelton, & Yaffee, 1994). For example, Martin (1994) compared two patients, E.A., who demonstrated a short-term memory deficit on phonologically-based tasks such as a rhyme probe task, but normal sensitivity to semantic content as measured by word and non-word retention task and a category probe task; and A.B., who showed the reverse pattern on a sentence processing task requiring lexical semantic storage. Martin compared error rates and reaction times in a sentence anomaly task where subjects were presented acceptable and unacceptable sentences such as those in Example 1.11-Example 1.13 and asked to press a button as soon as the sentence stopped making sense.

Example 1.11: The *fluffy* shriek came out of the room.

Example 1.12: The *fluffy*, *surprised* shriek came out of the room.

Example 1.13: The *fluffy*, *small*, *surprised* shriek came out of the room.

These sentences required subjects to maintain an increasing number of unintegrated adjectives (or nouns in similar sentences with conjoined subjects) in memory until they were able to fully interpret them by integrating them with their grammatical head. The import of lexical-semantics in this task becomes clear when one realizes that it is the integration of items which determines their precise meaning, for example *dry* has a different meaning if it refers to a towel, versus when it refers to a martini. (cf. Kintsch (2001) for further discussion). If patients suffered from a deficit specifically related to the retention of word meanings (i.e. lexical-semantics), then this task is predicted to be hard, and critically, harder than a task which permits immediate integration, such as the processing of sentences like those in Example 1.14-Example 1.16.

Example 1.14: The children played in the water that was *dry* until they got tired.

Example 1.15: The children played in the water that was *cold* and *dry* until they got tired.

Example 1.16: The children played in the water that was *cold*, *blue*, and *dry* until they got tired.

On the other hand, if the difficulty is simply related to retention of any type of semantic information, then the point at which integration occurs should make no difference. In addition, if the difficulty was related to a general short-term memory deficit and not a specific lexical-semantic deficit, then A.B. is predicted to perform as well as E.A. Martin found that A.B. did in fact have much greater difficulty with the unintegrated conditions than the integrated ones as compared with E.A. and with normal control subjects. Moreover, in a separate test of A.B's ability to maintain incomplete syntactic structure, Martin asked him to judge the grammaticality of sentences such as "Did the exterminator already sprayed harmful pesticides?" The patient A.B. showed no difficulty with making these judgments, and since the ungrammaticality did not occur until late in the sentence, she concluded that the earlier findings were not related to difficulties with syntactic storage. Processing dissociations with these patients, as well as others (cf. (Martin & Freedman, in press; Martin, Lesch, & Bartha, 1999)), have led Martin to suggest a further fractionation of working memory into separate phonological, syntactic, and lexical-semantic components. Moreover, these data suggest that Waters and Caplan's "encapsulation" of syntactic processing from lexical-semantic processing of the sort required in the Ferreira and Clifton (1986) and Trueswell et al. (1994) studies may in fact be correct.

1.2.4 Lewis' Type-specific Storage

The proposals we have seen thus far have been attempts to characterize the working memory capacity that supports language processing, together with how limits on that capacity account for language comprehension breakdown. They have done so using broad strokes, however, and often with one foot firmly planted in the world of neuroscience, and its desire to map out neurological correlates of language processing. Lewis (1996) takes a different approach, focusing on the representational language used for *encoding* intermediate representations of sentence processing, and demonstrating how storage of items with similar codes can produce processing breakdown. In so doing, Lewis attempts to unify a basic property of linguistic processing, that multiple center-embeddings are difficult, with a basic property of memory, that of similarity-based interference. Lewis (1996) presents a review of "working memory types" which are subject to type-specific interference, including the well known work of Baddeley (1966) and Conrad (1963) who show that recall of phonologically similar lists of words, consonants, and nonsense trigrams is worse than dissimilar lists, and the work of Baddeley and Hitch (1974) and Logie, Zucco, & Baddeley (1990) on verbal versus visual-spatial encoding. He also suggests there is evidence for distinct, and interfering, codes for kinesthetic memory (Williams, Beaver, Spence, & Rundell, 1969), odor (Walker & John, 1984), and sign language (Poizner, Bellugi, & Tweney, 1981).

Unlike the Waters and Caplan and Martin proposals, Lewis has fallen short of suggesting that there are distinct working memory *capacities* for each of these types of processing, but his view is very much in their tradition. Instead, he suggests that when items are similarly *encoded*, interference arises from storing multiple similar items and processing breakdown occurs. For example, consider an extreme version of the complex object relative clauses we discussed in Section 1.3.1, given in Example 1.17.

Example 1.17: The boy that the man that the woman hired hated cried. (Lewis, 1996)

The structure of this sentence requires the storage of three noun phrases in the role of grammatical subject before they can be assigned to their verbs: *boy, man,* and *woman.* Lewis marshals other data suggesting a limit of two (or three) on verbal short-term memory (e.g., (Daneman & Carpenter, 1980; Gibson, 1991; Simon & Zhang, 1985)) to suggest that Example 1.17 is difficult because three similarly indexed items creates more interference than the language processor can manage, resulting in processing breakdown. In contrast, the easier to process subject relative clauses in Example 1.18 and Example 1.19 do not create high amounts of syntactic interference, because only two subject NPs, *boy* and *woman* in Example 1.18 and *boy* and *man* in Example 1.19 must be buffered during processing.

Example 1.18: The boy that hired the man that the woman hated cried.

Example 1.19: The boy that the man who hired the woman hated cried.

Notice that the Lewis proposal is actually just a way of creating the same sort of syntactic processing specialization proposed by Waters and Caplan, without a claim with respect to whether the actual processing resource pool is syntax specific. Since processing breakdown is defined as interference from similarly encoded items, this means that regardless of the encapsulation of the resource pool, the determinants of *syntactic* breakdown certainly *must* be syntactic (i.e. the syntactic coding itself).

1.3 Evaluating Storage vs. Retrieval

The views presented here provide a representative survey of how working memory has been conceived in its role as a support for language processing, and also as a source of limitation. As noted in the introduction to this section, however, these views have all focused on the *storage* aspect of working memory. In this work, I will attempt to make the case that it is rather the *retrieval* aspect of working memory that provides more explanatory power in an account of processing breakdown. For example, there is actually very little data in support of the storage account of complexity effects in the classic subject versus object relative clauses presented in Example 1.1 and Example 1.2, repeated here as Example 1.20 and Example 1.21, with each word numbered.

Example 1.20: The reporter that attacked the senator admitted the error.

1 2 3 4 5 6 7 8 9

Example 1.21: The reporter that the senator attacked admitted the error. (King & Just, 1991)

1 2 3 4 5 6 7 8 9

While this comparison has long been the focus of models attempting to account for complexity effects, there have been surprisingly few empirical studies which directly compare processing of these structures with online measures, and of these even fewer permit comparisons in regions other than that containing words 6 and 7 in the examples above (i.e. the relative clause/main clause transition) (cf. Table 1.1 Table 1.2). Measures of the region prior to word 6 are critical for determining whether difficulty is associated with storage versus retrieval because the sentences permit detailed predictions of when each is expected to play its role. For example, in the subject relative in Example 1.20, the storage effect begins from the beginning of the relative clause *that* (word 3) because the NP *the reporter* must be buffered until it can be integrated with its verb. Thus, storage should begin here and proceed until the final retrieval (words 3-6). Retrieval effects in this sentence may occur immediately after the *that* when the subject of the relative clause verb must be identified (word 4), and at the main verb *admitted* when the subject of this verb must be identified (word 7). In contrast, the storage effects for the object relative sentence in Example 1.21 should also begin at the *that* (word 3), and continue until the main verb (words 3-6). However, this effect is likely to be greater than in the subject-relative clause because two unintegrated NPs must be maintained until the embedded verb occurs (word 6). This means that at least at word 5, the object relative clause should be more difficult than the subject relative clause if the source of the difficulty is due to storage. Moreover, if storage is the only source of differences between these sentences, then processing at word 7 should actually be similar for the object relatives and the subject relatives, because once the embedded verb has occurred in word 6, the increased storage load in object relatives is relieved. If, on the other hand, the difficulty is due to retrieval, then processing of the object relative should be identical to that for the subject relative up until the two verbs (words 6-7), where retrievals must occur in order to integrate the nouns with their verbs. Thus, if the source of the difficulty is due to retrieval, difficulties for the object-relative should appear at words 6-7, where retrievals must occur, and critically, reading times should increase earlier than for the subject relative, since the first retrieval occurs only at word 7 in subject relatives. In sum, the predictions of the two accounts are as follows: according to the storage account, the two sentences should diverge on words 5-6, becoming similar again at word 7 because the extra NP has been integrated with its verb; according to the retrieval account, the two sentences should not diverge until words 6-7, both of which should be more difficult for the object relative .

Table 1.1: Studies of Subject vs. Object relative clauses with online measures.
These studies do not differentiate reading times in the regions prior to words 6 and 7 (the relative clause/main clause transition).

Study	Stimuli	Paradigm	Dependent measures	Results	Comments
(King & Just, 1991), Experiment 1	SR: The / reporter that attacked the/ senator / admitted / the error. OR:The / reporter that the senator / attacked / admitted / the error	Procedure: Word-by-word moving window Subjects: Two groups split on reading span using (Daneman & Carpenter, 1980).	Raw reading time in each region (defined with slashes in Stimuli column).	Retrieval region: For comprehending subjects only: Word 6: OR > SR, Word 7: OR > SR. Storage region: No statistics reported but means and standard errors for the initial region suggest no differences for either memory span group.	Evidence of effects in the retrieval regions. No evidence of effects in the storage region.
(Stine-Morrow, Ryan, & Leonard, 2000)	SR: The pilot that the nurse / admired / dominated / the conversation / etc. OR: The pilot that admired the / nurse / dominated / the conversation / etc.	Procedure: Word-by-word presentation, but unstated if cumulative. Subjects: 2 age groups: Old (M = 70.9 years) and Young (M = 19.5 years).	Read times in regions (defined with slashes in Stimuli column).	Retrieval region: No statistics presented, but graph of data suggest: Word 6: OR > SR for young only, Word 7: OR > SR for young only. Storage region: No apparent sentence type differences in initial region.	Evidence of effects in the retrieval regions. No evidence of effects in the storage region.
(Booth, MacWhinney, & Harasaki, 2000), Experiment 1	SR: The / principal that tripped the / janitor / used / the phone / to call / home. OR: The / man that the captain / invited / built / the stage / for the / band.	Procedure: Word-by-word moving window. Subjects: Children aged 8-11.	Mean read times in regions (defined with slashes in Stimuli column)	Retrieval region: Word 6: OR > SR, Word 7: OR > SR Storage region: No differences in initial region..	Evidence of effects in the retrieval regions. No evidence of effects in the storage region.
(Booth et al., 2000), Experiment 2	Same as above	Auditory moving window, otherwise same as above.	Mean listening time in the regions (defined in Stimuli column).	Retrieval region: Word 6: OR > SR Word 7: OR > SR Storage region: No differences in initial region.	Evidence of effects in the retrieval regions. No evidence of effects in the storage region.

Table 1.2: Studies of Subject vs. Object relative clauses in pre-retrieval region.
These studies differentiate words prior to position 6, allowing clear evaluation of effects in the storage region.

Study	Stimuli	Paradigm	Dependent Measures	Results	Comments
(Wanner & Maratsos, 1978)	SR: The witch who despised (1) sorcerers frightened (2) little children. OR: The witch whom sorcerers (1) despised frightened (2) little children	<u>Procedure:</u> Word-by-word reading with interpolated recall task at interruption points 1 and 2 above. Comprehension questions at end. <u>Subjects</u>: Harvard/Radcliffe students	Accuracy on the two tasks summed.	<u>Retrieval region:</u> No difference at interruption point 2 <u>Storage region:</u> OR > SR at interruption point 1	Statistics reported only for the sum of the two task results. Lack of retrieval effects may be due to summation over tasks: differences appeared in the question task, but not in recall. Differences in the storage region.
(Holmes & O'Regan, 1981)	French. SR: Je crois qui le sauvage qui / va attaquer / le chasseur/ mont sur un cheval noir. OR: Je crois que le sauvage que / la chasseur / va attaquer / monte sur un cheval noir. TOR: Je crois que le sauvage que / va attaquer / le chasseur / monte sur un cheval noir. *** TOR is a transposed Object relative, with object relative meaning but SR surface order.	<u>Procedure:</u> Eye-movements <u>Subjects:</u> French native speakers	Initial fixations in regions defined by slashes in Stimuli column. Although there are different numbers of words in these relative clauses, I'll continue to refer to the embedded verb as word 6 and the main verb as word 7 in reporting the results.	<u>Retrieval region:</u> Word 6: OR > SR, TOR Word 7 (aggregated with following words): marginal difference. <u>Storage region:</u> Word 5: SR, TOR > OR	Evidence of effects in the retrieval region for Word 6. Cannot interpret Word 7 effect because it is aggregated. Storage hypothesis contradicted because SR & TOR > OR. Similar TOR and SR results argue against storage, since TOR doesn't require storage.
(Ford, 1983)	SR: The reporter that attacked the senator admitted the mistake. OR: The reporter that the senator attacked admitted the mistake.	<u>Procedure:</u> Word by word moving window with lexical decision task on each word. <u>Subjects</u>: Stanford students	Decision times at each word	<u>Retrieval region:</u> Word 6: OR > SR Word 7: OR > SR Word 8: OR > SR. <u>Storage region:</u> Words 2-5: no differences	Evidence of effects in the retrieval region with spillover to Word 8. No evidence of effects in storage region.

Table 2 (continued)

Study	Stimuli	Paradigm	Dependent Measures	Results	Comments
(King & Kutas, 1995)	SR: The reporter who harshly attacked the senator admitted the error. OR: The reporter who the senator harshly attacked admitted the error.	Procedure: ERPs with comprehension probes. Subjects: UCSD students, separated into good/poor comprehenders based on median split to comprehension probes scores.	Single word ERPs	Retrieval region: N2/V1: OR more negative (LAN) than SR. Main verb: OR more negative (LAN) than SR. Effect is more pronounced for good comprehenders. Storage region: V1/N2: no difference Beginning of relative clause shows a LAN for ORs that is more pronounced for poor comprehenders.	Evidence supports effects in retrieval regions, specific to verbs in those regions. No support for storage except at beginning of relative clause.
(Gordon, Hendrick, & Johnson, in press), Experiment 1	SR: The cook that helped the plumber quit work after a month. OR: The cook that the plumber helped quit work after a month.	Procedure: Self-paced RSVP with comprehension Qs Subjects: UNC students	Raw reading times.	Retrieval region: Word 6: OR > SR Word 7: OR > SR Storage region: Words 3-5: no differences	Evidence of effects in the retrieval regions. No evidence of effects in the storage region.
(Gordon et al., in press), Experiment 2	Same as above, crossed with "you" as relative clause NP instead of "the plumber"	Same as above.	Same as above	Retrieval region: Words 6&7: OR> SR only when "the plumber" is the intermediate NP. The difference disappears if "you" is the intermediate NP. Storage region: Words 3-5: no differences	Evidence of effects in retrieval regions, but only when similarity of intervening items is high. No evidence of effects in storage region.
(Gordon et al., in press), Experiment 3	Same as above, Exp 2, but with proper names instead of "you"	Same as above	Same as above	Same as above	Same as above

The data available from the studies summarized in Table 1.2 show only one study with significant differences in the region where the storage effect occurs (i.e. word 5), this being the study by Wanner and Maratsos (1978). This study is widely cited in support of storage accounts, as it was the first to formalize the storage explanation, which it did using in an ATN (augmented transition network) model of sentence processing. They presented the HOLD hypothesis, which suggests that processing of relative clauses involves postponing a decision about the grammatical function of a noun phrase (i.e. adding it onto a HOLD buffer) until a linguistic gap identifying its role is encountered. They argued that this HOLD necessarily places an increased load on memory from the beginning of the relative clause until the point at which the gap is discovered and the NP is removed from the HOLD list, i.e. words 3-5 as discussed above. Consequently, this is the *only* study to recognize this region as a critical region for hypothesis testing.

Their results do appear to support the HOLD hypothesis, however, there are numerous questions with the procedure used to obtain them. First, they employ no direct measure of reading times, but rather present analyses on combined error scores from two tasks, one occurring during sentence processing itself, and one occurring after the sentence has been read. The online task occurred at two interruption points during the sentence, one in the storage region and one in the retrieval region, and required subjects to memorize a list of names before continuing to read the target sentence. At the conclusion of the sentence, subjects were asked to recall the names on the list, and then to answer a comprehension question about the sentence, and it is the error scores for these two tasks that constituted the Wanner and Maratsos data. Thus, rather than a measure of subjects' linguistic processing of these two sentences, their data is rather a measure of subjects' recall ability in an interfering task situation, making conclusions regarding the source of sentence complexity effects difficult to defend.

A second study that shows some support for the storage account because of observed differences in the storage region is the ERP study of King and Kutas (1995). While analyses of the single-word ERPs at word 5 in the storage region show no differences, they did observe a LAN component for object relative clauses at the beginning of the relative clause (i.e. the adverb in SRs and the article in ORs, see Table 1.2 for stimuli). This provides some evidence that the grammatical difference between the relative clauses does invoke a processing response, especially in light of the fact that the LAN was more pronounced for poor comprehenders. However, since this effect disappeared before the occurrence of the actual NP being stored, it seems unlikely that this is a response to the *storage* operation per se. Rather, King and Kutas suggest that the response is due to decreased expectations, especially among poor comprehenders, that the object relative structure would occur.

Thus, in the few studies to find differences in the storage region we do not find clear support for the storage account of processing differences between subject and object relatives. Rather, the data appear to fit a retrieval account much more clearly, with nearly all studies in Table 1.1 and Table 1.2 observing differences in the retrieval area (words 6 & 7). One particularly interesting study was conducted by Holmes and O'Regan (1981), who looked at initial-fixations in an eye-tracking study using subject and object relative clauses in French. They observed differences in the storage region (word 5) in the opposite direction to that predicted by the storage account, in that subject relative clauses were actually more difficult than object relative clauses at this point in processing. These results can be explained by a retrieval account; however, since although this is the point where storage must occur in the object relatives, it is the point where the retrieval of the subject of the relative clause occurs in the subject relatives. Thus if additional processing cost is associated with retrievals, then we would expect the subject relative to be more difficult in this area. Moreover, the study also included a French transposed object relative (TOR) construction, which permits preserving subject relative clause surface order with object relative clause meaning (see Table 1.2 for stimuli). Thus, although the embedded verb occurs directly following the relative pronoun, the form of this pronoun signals that the clause is an object relative, and consequently the head of the relative clause should not be made the subject of this verb, but rather its object. These object relative structures do not require storage at the word 5 region, but rather a retrieval of the verbal object, just as the retrieval of a verbal subject is required at word 5 in subject relatives. Not surprisingly then, these transposed object relatives were also more difficult than the object relatives at word 5, but identical in difficulty with the subject relatives. Then when retrieval was required for the object relatives at word 6, the pattern was reversed. Object relatives were fixated longer than either the subject relatives or the TOR sentences at this word, again fitting with the hypothesis that processing cost is associated with *retrieval* operations rather than storage.

It is important to be clear, however, that I am not suggesting that the storage operations that distinguish these two sentences are cost free, but merely that it is not this cost that creates the processing difficulty observed empirically. Data that clarify this point are available from an ERP study conducted by Fiebach, Schlesewsky, and Friederici (2001; submitted) with German WH-questions. While these data do not bear directly on the relative clauses we have been discussing, the WH-question constructions do involve the same structural characteristics, in that the object questions require buffering an element until its syntactic position is revealed later in the sentence by a phonologically null *gap*, indicated here by "____".[1]. Consider Example 1.22, which illustrates a subject WH-question and Example 1.23, which illustrates an object WH-question:

Example 1.22: Thomas fragt sich, wer am Mittwoch den Doktor verständigt hat.

Thomas asks himself, $\text{who}_{(NOM)}$ on Wednesday $\text{the}_{(ACC)}$ doctor called has.

Example 1.23: Thomas fragt sich, wen_i am Mittwoch der Doktor $\_\_\__i$ verständigt hat.

Thomas asks himself, $\text{who}_{(ACC)}$ on Wednesday $\text{the}_{(NOM)}$ doctor called has.

---

[1] The term *gap* has been used to refer to the syntactic position in which a constituent would appear in the declarative version of the same sentence, and from which this constituent is assumed to have *moved*. This movement creates a dependency chain from the moved item (referred to as the *filler*) and the *gap* (Chomsky, 1981). Thus, it is at the position of the gap that the correct structural interpretation of the filler can be made, and it is at this point that costs relating to the integration of the filler into the syntactic structure are expected to occur.

In Example 1.23, the appearance of the accusative marked pronoun *wen* indicates that an accusative marked item will occur downstream, cueing the parser to interpret the following structure in this context, and in particular tell it that it must buffer the relative pronoun for later structural placement. This storage does not occur in the subject-question of Example 1.22 because the nominative pronoun *wer* permits immediate structuring as a subject of the embedded clause. Thus, just as in the relative clauses discussed previously, structural interpretation of the object WH-question requires a storage operation early in the sentence that is not required in the subject WH-question, and is therefore an obvious candidate to explain data showing the object WH-question to be more difficult (e.g., (Bader & Meng, 1999; Faneslow, Kliegl, & Schlesewsky, 1999)). To test the storage hypothesis more directly, Fiebach et al. (2001) compared the sentences in Example 1.22 and Example 1.23 with those in Example 1.24 and Example 1.25, which require storing the unintegrated relative pronoun in the object WH-question for much longer.

Example 1.24:

Thomas fragt sich, wer am Mittwoch nachmittag nach dem Unfall den Doktor
Thomas asks himself, who(NOM) on Wednesday afternoon after the accident the(ACC) doctor
verständigt hat.
called has.

Example 1.25:

Thomas fragt sich, wen_i am Mittwoch nachmittag nach dem Unfall der Doktor $\_\_\__i$
Thomas asks himself, who(ACC) on Wednesday afternoon after the accident the(NOM) doctor
verständigt hat.

called has.

Multi-word ERPs from the onset of the wh-pronoun until the occurrence of the embedded verb revealed a left anterior negativity (LAN) for the long object WH-questions (e.g. Example 1.25), but not for the long subject WH-questions, supporting the hypothesis that storage operations are costly in these sentences. Importantly, however, they did not find any difference between the two short WH-questions in this region, which they explain by noting that although storage must occur in order to interpret the short object WH-question, working memory can be freed earlier than in the long questions and so storage contributes little to the overall processing cost. A main effect of structural type did occur at the position of the gap in object WH-questions, with the occurrence of a P600, and was present for long and short sentences alike. Fiebach et al. follow previous interpretations of the P600 (e.g., (Kaan, Harris, Gibson, & Holcomb, 2000)) in suggesting its presence marks difficulty with the syntactic integration that must occur at this position, an integration that requires retrieval of the WH-pronoun in both long and short versions.[2] Thus, like the evidence available on relative clauses, this evidence suggests that the source of difficulty in object-extracted clauses is not due to storage, but rather due to the retrieval and integration of required constituents.

This focus on the retrieval aspect of processing costs provides an account of complexity effects in sentences with moved constituents which is very much in line with the body of research showing that gap-filling involves reactivation of the filler (e.g., (Hickok, 1993; Love & Swinney, 1996; Nicol & Pickering, 1993; Nicol & Swinney, 1989; Zurif, D., Prather, Solomon, & Bushell, 1993)). This research has shown that lexical items that are semantically related to the filler item are identified more quickly at the gap position than unrelated items. Crucially, however, there is no difference in identifying related or unrelated items at points prior to the gap, except for the point at which the filler item actually occurred, suggesting that the initial activation of the filler decays during the intervening material until a point in the sentence when it becomes reactivated, or *retrieved*, so that syntactic integration can occur.

There is some controversy, however, over whether this reactivation is mediated by the presence of the gap, or the phonologically null trace that generative linguistic theories assume to be present at the gap (cf.

[2] A natural question arising from this result is why the retrieval for the long object WH-question would not be more difficult than that for the short object WH-question, since the relative pronoun has had more time to decay. We will address this issue in section 1.5 and in 1.6 below.

(Chomsky, 1981)), as several studies have shown evidence of reactivation prior to the gap (e.g., (Boland, Tanenhaus, Garnsey, & Carlson, 1995; Traxler & Pickering, 1996); but see (Gorrell, 1993) for a discussion of difficulties with this account.) For this reason, the structures I will examine in the present study are *not* filler-gap structures, but rather those requiring a long distance attachment, as in the example in Example 1.26.

Example 1.26: The secretary knew that the woman who was sitting in the hallway was waiting for the chairman.

This structure is linguistically complex because the embedded relative clause separates the subject *the woman* from its verb phrase *was waiting*, requiring its retrieval in order to complete the long distance attachment. This allows a clear prediction of where retrieval effects will occur (i.e. at the verb *was waiting*) because there are no other points in the sentence where the dependency can be anticipated or completed. If, on the other hand, the complexity effects are due to *storage* difficulties of the intervening items, then those effects should occur in the intervening region prior to the phrase *was waiting*.

In addition to evaluating the storage vs. retrieval accounts on the basis of *where* effects occur, I will evaluate this distinction through manipulating two properties of the intervening region that are predicted to differentially affect retrieval and storage. The first of these, examined in Chapter 2, is the semantic properties of the intervening NP, which on a storage account should be irrelevant because the NP must be stored regardless of its meaning. The second, examined in Chapter 3, is the referential status of the intervening NP as either "old" or "new" vis-a-vis the current discourse. A storage account would predict that old items are not costly because they are already stored, taking up no additional resources. The retrieval account predicts that old items may be costly if they are too old, as they will become more difficult to retrieve the more they have decayed.

In order to understand why semantic and referential properties are predicted to affect retrieval, a precise account of the retrieval mechanism is necessary. I will assume that retrieval in sentence parsing is an instance of the associative retrieval described in the episodic memory retrieval literature, reviewed in the next section. Heretofore, models of episodic memory have not been directly applied to the task of sentence processing; however, if complexity effects stem from retrieval difficulties, and not storage difficulties, then the same properties known to effect retrieval in this empirical work should also affect retrieval, and hence comprehension, of the sentence structures examined here.

1.4 Cue-based Memory Retrieval

Theoretical and empirical research on human memory has led to the development of a number of precisely specified models of episodic retrieval, for example Murdock's (1982; 1983) theory of distributed associative memory (TODAM),. Pike's (1984) Matrix model, Hintzman's (1984; 1988) MINERVA 2 model, and Shiffrin and colleague's (Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1981b; Shiffrin, Ratcliff, & Clark, 1990) search of associative memory (SAM) model (see Clark and Gronlund (1996) for a comparative review). Having been developed to account for evidence garnered via list learning in the encoding/retrieval paradigm (cf. (Tulving, 1983)), these models share the assumption that retrieval success depends on how well the cues at retrieval match the encoding conditions present when items were stored. Substantial evidence supports this assumption, for example, Thomson and Tulving (1970) presented subjects with a list of target words (e.g., FLOWER) in two encoding conditions: either in isolation, or with weak associates (e.g., fruit-FLOWER). They then manipulated the retrieval conditions so that subjects were asked to recall the targets with no cues, with the weak associates the words were presented with, or with strong associates they had not seen (e.g., BLOOM). They found that when words were studied in isolation, strong associates presented as recall cues resulted in a significant increase in the proportion of recalled words. When words were studied with weak associates, however, it was only these weak associates that assisted recall, and critically *not* the strong associates, which actually leads to *worse* recall than when the words are recalled without any cues. This result illustrated the importance of *context* cues for retrieval, and led to the proposal of the *encoding specificity* hypothesis, which suggests that retrieval is successful only to the extent that information extracted from the cues in the testing condition matches the way information was encoded at storage (Tulving & Thomson, 1973).

In this dissertation, I use the fact that retrieval depends on how well retrieval cues match storage cues as a framework for making predictions about when retrieval effects will occur in sentence processing. I will assume that each word in a sentence enters working memory, creating a word list just like those used in the list learning paradigms. Support for this type of representation in sentence processing was provided by Gibson and Thomas (1999) in a study examining the "missing VP" effect, where sentences such as Example 1.28 taken from Gibson and Thomas (1999), are perceived to be grammatical despite missing a verb for the matrix subject.

Example 1.27: The patient who_i the nurse who_j the clinic had hired e_j admitted e_i met Jack.

Example 1.28: *The patient who_i the nurse who_j the clinic had hired e_j met Jack.

Example 1.29: *The patient who_i the nurse who_j the clinic had hired e_j was smiling met Jack.

Example 1.30: *The patient who_i the nurse who_j the clinic had hired e_j for e_i met Jack.

It is only sentences which do not preserve any of the grammatical dependencies for the "forgotten" item which are perceived as correct, as illustrated by the unacceptability of Example 1.29 which preserves the dependency between *nurse* and *was smiling* while forgetting the expectation of a empty-NP for the relative clauses headed by *patient*. Similarly, Example 1.30 is perceived as ungrammatical despite preserving the empty-NP site by adding a preposition, and thereby satisfying the expectation of a grammatical dependency between the NP *patient* and a later constituent. Gibson and Thomas concluded that the memory representation underlying structure building is of the lexical items themselves, just as if the words were in a word list, since it is only in this case where forgetting would cause the loss of all associated syntactic predictions (i.e., if the representation included grammatical dependencies, then associated syntactic predictions would also be preserved, allowing sentences like in Example 1.29 and Example 1.30 to be perceived as correct.)

Sentence parsing is, of course, different from list learning. For example, in paired recognition of list items, one of the paired words is presented as a prompt for retrieving the word studied with it. This word serves as a context cue for the retrieval, which is a non-automatic operation imposed by the experimental procedure, and therefore extrinsic to the memory task itself. In addition, the words in a sentence that would constitute the "list" to be learned include naturally occurring dependencies that do not exist in typical paired recognition paradigms. It is precisely these dependencies which transform sentence parsing into an instance of paired recognition, as dependencies in the sentence prompt an automatic retrieval operation. Moreover, since words in the sentence function in particular grammatical roles because of the way they are used in that sentence, the grammatical structure of the sentence provides context cues for identifying the appropriate arguments for the verbs. Thus, in Example 1.31 when the verb phrase *was standing* is encountered, it contains a retrieval cue for a required subject, which can be identified by the grammatical roles assigned to items already parsed in the sentence.

Example 1.31: The secretary forgot that the student who was waiting for the exam was standing in the hallway.

Figure 1.1 illustrates the cues prompting the subject retrieval in this sentence. The verb has three retrieval cues associated with it, $\{N_{was2}, sing_{was2}, subj_{was2}\}$, which will match with both the NP *student* and the NP *who*, and require a process of differentiation, which will be easy since the NP who is an empty element.[3]

---

[3] This is a grammatically simplified presentation. In most analyses, the relative pronoun *who* will occur in complementizer position, and an empty trace element occurs as subject position of the relative clause. We have not represented the sentences here in this fashion so as to simplify the discussion. Note however that such a representation would make discrimination of the matched items even easier, since the complementizer would not match the retrieval cue, being of category C. Discrimination may also be aided by a specification of a non-null subject for the verb, thus eliminating the trace element from the match entirely. Consequently, the exact specification of the underlying grammatical representation is less important than the point that the target subject *student* is easily discriminated from the relative clause subject.

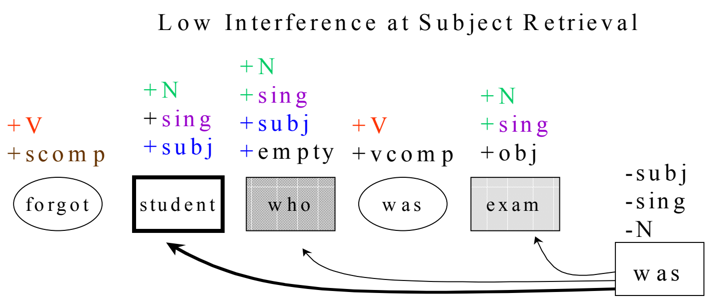


Figure 1.1: Cue-driven subject retrieval
for the sentence *The secretary forgot that the student who was waiting for the exam was standing in the hallway.*

Cue-based retrieval models predict that processing will proceed smoothly when the available retrieval cues identify the required dependents unambiguously, as occurs in Example 1.31. This is not the case in Example 1.32 because not only is *student* a subject, but so is *exam*:

Example 1.32: The secretary forgot the student who knew that the exam was important was standing in the hallway.

Differentiation is much more difficult in this sentence, and the cue-based retrieval models would predict that this sentence is difficult to process, which of course it is. In fact, this is precisely the type of sentence that led Lewis (1996) to suggest that processing difficulty was due to storage interference from having to temporarily store three subject NPs. The focus on retrieval cues preserves this argument to some extent, as difficulty is related to the fact that the three NPs are stored, but the account proposed here does not suggest that it is storage *per se* which causes the processing breakdown, but rather the inability to unambiguously select the appropriate item from competing candidates. Thus, while the retrieval models can account for the same facts as the storage model of Lewis (1996), effects associated with context cues, as defined here, can not differentiate the retrieval and storage accounts. For this we turn to two additional types of cues, interitem and self-strength cues.

Interitem cues come from pre-existing semantic associations between items in a list. Their role in retrieval has been demonstrated empirically in studies showing that the probability of successfully recognizing (and hence retrieving) an item decreases as the size of the list increases (e.g., (Gillund & Shiffrin, 1984; Ratcliff & Murdock, 1976)), and also if the similarity of the items in the list increases (e.g., (Bowles & Glanzer, 1983; Hintzman, 1988)). For example, Hintzman (1988) selected 288 familiar nouns distributed across 48 semantic categories with high within-category similarity and low between-category similarity, (e.g., booklet, pamphlet, periodical, Scotch, rum, brandy, minister, priest, rabbi, etc.) The 48 categories were divided into 4 sets of twelve, and each set was assigned a presentation frequency of 0,1,3, or 5 which determined how many items from that category would be presented to the subjects. During the study-phase, subjects saw these words, together with 92 unrelated filler items, and were asked to rate them as to how much activity they involve. In the retrieval phase, subjects were presented with a test-list and were asked to identify the words that had appeared during the study phase (i.e., recognition test). Hintzman found that hit rates and false alarm rates both increased monotonically as the number of same-category items increased, showing that a subjects' tendency to call an item *old* increased with the number of similar items in the list, regardless of whether the item itself was actually new or old. This finding shows clearly that retrieval is also influenced by the other items in memory, and especially the associative strength between other items and the test item itself.

Extrapolating this result onto the case of sentence parsing leads to the prediction that semantic associations between items in a sentence can lead to retrieval difficulty. To illustrate, consider an additional semantic dimension in the retrieval cues supplied by *was standing*, in Example 1.32, repeated below, creating a retrieval probe with the cues, $\{N_{was2}, sing_{was2}, subj_{was2,}\ standable\}$.

Example 1.32: The secretary forgot the student who knew that the exam was important was standing in the hallway.

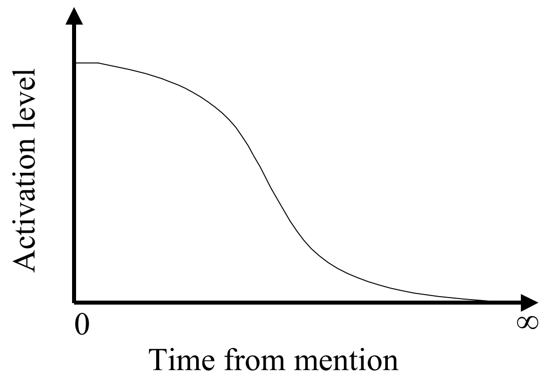
In this example, the intervening NP *the exam* fits the syntactic cues, but not the semantic ones because it does not make sense with the verb *stand*. If that NP were to fit with the semantic cues from the verb, as does the NP *professor*, in Example 1.33, then the strength of this association is predicted to make retrieval of the proper NP *the student* more difficult.

Example 1.33: The secretary forgot that the student who knew that the professor was important was standing in the hallway.

Notice, however, that a storage account would predict no difference between Example 1.32 and Example 1.33, since in both cases resources are required to store the intervening NP and their syntactic structure is identical. This prediction is tested empirically in Chapter 2.

A third factor found to affect retrieval is the strength, or activation, of a particular item in working memory, known as its *self-strength*. The initial activation of an item is a function of properties of the item itself, such as its frequency, as well as the conditions of its encoding, i.e., how long it was studied. As time passes, this activation decreases monotonically, as illustrated in Figure 1.2, unless it receives an additional activation boost from being repeated or reaccessed.

Figure 1.2: Activation function for discourse referents in a situation model



For example, Graf and Mandler (1984) found that in a task where subjects were required to rate their degree of "liking" a list of words, their subsequent recall of these words improved if the word occurred twice in the presentation list. In a follow-up study, they presented a different list of low-frequency words to be learned via the same encoding task and tested subjects both in a free recall task and in a completion task in which subjects were given the first three letters of a target item and were required to complete the item. They found that when the completion task followed the recall task, subjects' performance was significantly improved, but there was no improvement if the completion task occurred first. They argued that this was due to the retrieval of the item required in the recall task, which restores that item into working memory and affords an additional opportunity for rehearsing and encoding it in the new context, much as if the item had been repeated. This results in an increase in the self-strength of that item together with its association strengths with other items in the list, making it more accessible for the subsequent completion task. In contrast, the completion task does not afford the same recoding opportunity, so there is no resulting increment in the item's self-strength or additional elaboration of interitem associations and, consequently, no benefit for the recall task is observed.

A similar result was found by Raaijmakers and Shiffrin (1981a), who found that cued recall of paired associates was impaired by prior free recall of isolated words appearing in the same study list, but that the reverse effect was not found when cued recall occurred first. As in the Graf and Mandler study, they suggested that this effect could be explained by increases in the context, interitem, and self-strengths of the item after successful retrieval. These increases result in more recall overall, primarily due to the stronger interitem associations, which create a kind of *retrieval plan* for the recall task (Raaijmakers & Shiffrin, 1981b). In contrast, when paired associates are retrieved, the increase in interitem strength is not as great since the cue for that item (i.e. part of the pair) is not associated with any other items in the set. For example, in a list containing (A, B, XY, C, DE, FG), the item A is associated with all the other items in the list via context and interitem associations, but X, a retrieval cue of Y, does not have further associates in the list because it was not studied individually (Humphreys, 1976; Tulving & Thomson, 1971). This results in poorer performance in the paired recall task because A can cue more items than X. However, when the paired recall task occurs prior to the free recall task, the advantage due to interitem associations is lost because the previous retrievals strengthen the pair disproportionately. The probability of sampling that item in future retrievals is thus increased, and results in fewer new items being recalled (Gillund & Shiffrin, 1984). This interference can be overcome when cued recall occurs following free recall because the cues provide direct access to studied items, but when cued recall occurs first, no mechanism is available to overcome the interference generated by the task.

In sentence processing, the effect of self-strength can be seen in two areas. First, items which have occurred more recently have a higher self strength, creating interference for retrieving less recent items, even if they do not completely match the available retrieval cues. Figure 1.3 illustrates how the self-strength, or activation, is independent of the number of directly matching cues, but can nevertheless affect how easily an item with more matching cues is retrieved. If a less perfectly matching item happens to have a higher self-strength, then the probability of matching this item in the retrieval becomes similar to the probability of matching an item with more directly matching cues, and retrieval becomes more difficult because the highly active, but less matching item competes with the better matching item.

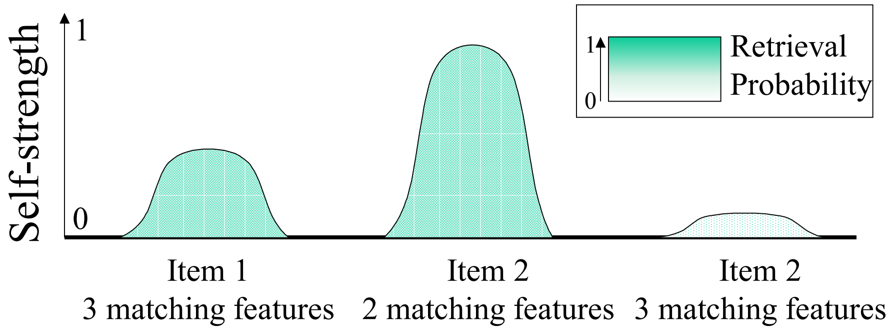


Figure 1.3: High self-strength varies independently of the number of matching features.

This explains why interference from *professor* in Example 1.34 is predicted, despite the fact that it serves in the grammatical role of object of the preposition, and hence does not match the syntactic cues of the verbal *was standing.* This prediction will be tested in Chapter 2.

Example 1.34: The secretary forgot that the student who was waiting for the professor was standing in the hallway.

A second approach to investigating effects of self-strength in sentence processing involves the referential status of discourse referents as *known* in the discourse or *new* to the discourse, meaning that the current mention is their first mention. Items which are known in the discourse, but which must be reactivated in order to interpret their meaning, are predicted to be more costly than items being mentioned for the first time, which will already be highly active. In full NP.

Example 1.35, this means that the NP *she* will make the sentence more difficult than if it were replaced by a full NP.

Example 1.35: The secretary forgot that the boy who knew that she was busy was waiting in the hallway.

In contrast, a storage account does not predict this difference, since the fact that the referent has already been stored means that no additional decrements to working memory capacity are required in order to process the pronoun. Rather, the sentence with the full NP is actually predicted to be more difficult than the one with the pronoun because in that case three NPs must be stored instead of merely two. I will provide an empirical test of prediction and other related predictions in Chapter 3.

1.5 Overview of Experiments

In the following chapters I present five experiments whose results illustrate the role of retrieval in sentence processing. In each case, the variables being manipulated are either not predicted to affect processing according to a storage account, or else this account predicts an effect in the opposite direction compared with predictions of the retrieval account. Experiments 1 and 2 in Chapter 2 evaluate the effect of semantic properties of NPs in the region intervening between two grammatically dependent items. The first experiment provides an initial evaluation of whether the semantic effect exists using an offline dependent measure, while the second provides online data that affords the opportunity of localizing the effect in either the storage or retrieval regions of the sentence. Experiments 3 and 4 in Chapter 3 evaluate the effect of the referential status of intervening NPs, using pronominal NPs to refer to already established referents. Experiment 5 provides online evidence illustrating the role of intermediate retrievals during sentence processing, and suggesting that these retrievals have an inverse effect on the availability of other items in the situation model, making it more difficult to integrate the new grammatical dependency in the broader context of the sentence. Chapter 4 attempts to integrate these results in a two stage model of sentence processing in which the retrieval mechanism takes over much of the work normally assigned to the syntactic parsing mechanism, and serves as the foundation for the construction of a broader representation of sentence meaning.

2 Syntactic and Semantic Interference

2.1 Introduction

Taking a cue from the episodic memory retrieval literature, in Chapter 1 I suggested that the successful creation of grammatical dependencies required the ability to unambiguously identify the correct dependent based on retrieval cues provided by the grammatical head. In the current chapter, I illustrate the role of syntactic and semantic retrieval cues provided by the grammatical head, with special attention to how these cues combine in order to select a single item to fill a grammatical dependency. In order to make this discussion concrete, I will present it in terms of the search of associative memory (SAM) model proposed by Shiffrin and colleagues (Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1981b). This model is particularly appropriate for this discussion because it differentiates retrieval cues along the same lines investigated here (i.e., contextual, semantic, and self-strength). Accordingly, I will first describe the model and then explain the cue combining effects it predicts, including effects of syntactic and semantic interference. Following this I present the empirical work evaluating these effects in the context of sentence processing.

2.2 Search of Associative Memory (SAM) model

The search of associative memory (SAM) model explicitly distinguishes three types of retrieval cues: context cues are assumed to have strength *a* to items encoded in a particular study situation (e.g., a study list or a study environment), with the context cue to the current situation being greater than that to any previous contexts so that items learned in those contexts are effectively ignored. In addition, an item has cue strength *b* to any images with which it is associated, either via rehearsal in the current context or via pre-existing semantic associations, and an item's self-strength *c* is the pre-existing strength of the item in working memory. The model also includes the cue strength *d* to account for residual associations between items which were not rehearsed together, or which are not associated directly in the current context. For the purposes of the current work, only cue types *a*, *b*, and *c* are relevant. In the realm of sentence processing, context cues, type *a*, refer to the syntactic role a particular item holds in the sentence being processed. Interitem cues, type *b*, represent the semantic congruency between the retrieval probe and a candidate filler, that is, how well a particular word fits as the filler of the particular grammatical role required by the grammatical head. Self-strength cues, type *c*, refer to the amount of activation the item currently has in working memory. According to the SAM model, these cue types are combined in a *retrieval structure*, which gives the strength of the relationship between each possible probe cue (Q_1, ..., Q_m) and the memory trace (I_1, ... I_n) for each item in the rehearsed list. For example, the retrieval structure for list L, containing pairs AB and CD, is presented in Figure 2.1.

Memory traces						
			I_1	I_2	I_3	I_4
			A	B	C	D
Cues	Q_1	A	c	b	d	d
	Q_2	B	b	c	d	d
	Q_3	C	d	d	c	b
	Q_4	D	d	d	b	c
	Q_5	Context	a	a	a	a

Figure 2.1: Sam retrieval structure following study of two word pairs, AB and CD in List L. Item cues A, B, C, and D and context cues are listed at the left-most column. Memory traces of the items A,B,C,D are in the top rows. Entries in the matrix are cue-to-item strengths.

The probability of retrieving a particular remembered item is a function of the strength of association S between the probe cues and its memory trace, denoted as $S(Q_j,I_i)^{w_j}$ where w_j denotes the relative saliency (e.g., items in the focus of attention) of the different cues. Specifically, Equation 2.1 specifies SAM's retrieval function:

$$P(I_i|Q_1, \ldots, Q_m) = \frac{\prod_{j=1}^{m} S(Q_j, I_j)^{w_j}}{\sum_{k=1}^{N} \prod_{j=1}^{m} S(Q_j, I_k)^{w_j}}$$

Equation 2.1: SAM retrieval function

It is interesting to point out that this relationship between the probe cues and the memory trace of an item is the same relationship used in the ACT-R (Anderson & Lebiere, 1998) model of cognition, in which the activation A of a memory chunk *i*, is given in Equation 2.2, where B_i is the base-level activation of the chunk, W_j is a property of the focus of attention, and S_{ji} which is the strengths of association between the elements of *j* and the chunk to be retrieved.

$$A_i = B_i + \sum_{k=1}^{j} W_j S_{ji}$$

Equation 2.2: ACT-R activation function

However, the SAM model allows a more explicit examination of the types of information affecting retrieval than does the ACT-R model because it distinguishes the cue types we have mentioned above and permits explicit predictions about how they interact. For example, given the retrieval matrix in Figure 2.1 and Equation 2.1, we can predict that the probability of retrieving item C given the proper context cues and the item D is *ab* divided by the overall activation of the items in the study list by this cue set, or $ac + ab + 2ad$. Similarly, the probability of retrieving item C given the same clues plus item C itself is *abc* divided by $2ad^2 + 2abc$. Thus we can see that the effect of adding C as a probe cue depends on the value of the self-strength of C, but also on how this item is related to other items in the memory set. In ACT-R, these interactions are all contained within the unitary term W_j and are therefore less straightforward to interpret outside the context of the specifications of a particular model.

Two key assumptions underlie the way the SAM model uses the different types of cues we have discussed. First, cue strengths are multiplicative, resulting in retrieval of the item that is most strongly associated to *all* the available cues. This is evident from the product in Equation 2.1, the result of which is to create a single joint probe of memory based on the associative strengths of all the cues available. Second, the retrieval depends on the strength of this integrated cue compared with the strengths of all other images in the memory set. This is seen in the summation term in the denominator of Equation 2.1. These two properties afford a search which is more properly characterized as *global matching*, and which is completely different from a serial search involving a systematic examination of individual memory items. The character of the global matching process, as well as a fact about global matching important for the predictions in Experiments 1 and 2, is best illustrated via example. Consider the retrieval matrix following the study of two word triples ABC and ABD, shown in Figure 2.2.

			Memory traces			
Cues			I_1	I_2	I_3	I_4
			A	B	C	D
	Q_1	A	c	2b	b	b
	Q_2	B	2b	c	b	b
	Q_3	C	b	b	c	d
	Q_4	D	b	b	d	c
	Q_5	Context	a	a	a	a

Figure 2.2: SAM retrieval matrix following study of two word triplets, ABC and ABD

Using Equation 2.1 above, the probability of retrieving item C given the cue A and the appropriate context cues is *ab/(ac + 4ab).* Notice that the probability of retrieving item D given the same cues is also *ab/(ac +4ab).* Since cue-based retrieval means that the entire matrix is probed using the available cue set, both of these items will become activated in response to the available cues. This sets up a response competition in which SAM has identified multiple matches to the cue set, and must sample from this active set in order to recover a particular item. Accordingly, the probability of recovering a particular item decreases inversely with the number of items matched by the retrieval cues. In this way, SAM suggests that interference effects result from the necessity of selecting among a set of retrieved items generated by insufficiently discriminating cues (Mensink & Raaijmakers, 1988). I will test this prediction directly in Experiments 1 and 2.

2.3 Interference effects

The goal of this chapter is to test the predictions of cue-based retrieval and global matching in the context of sentence processing. As discussed in Chapter 1, this is achieved through the use of syntactic structures requiring attachment across an intervening region that contains items that may or may not fit the retrieval cues provided by the head of the grammatical dependency. If parsing occurs using a cue-based retrieval mechanism, then intervening items that also match these cues will supply interference, increasing the difficulty of completing the long distance attachment. For example, in sentence Example 2.1 below, the noun phrase *the resident* must be attached as the subject of the verb phrase *was complaining* over the intervening region *who was living near the warehouse*.

Example 2.1: The social worker was surprised that the resident who was living near the dangerous warehouse was complaining about the investigation.

A cue-based retrieval is initiated by *was complaining*, which supplies retrieval cues to pick out *the resident* from the previously occurring word string. These cues require a singular noun which is a subject and which makes sense with the verb *complain* and in sentence Example 2.1 there are only two matches for this, *worker* and *resident*. Since *resident* is more recent it will have the highest probability of being retrieved as the subject of *was complaining*, and the items in the intervening region supply little interference for this selection.

In contrast, a cue-based retrieval mechanism will have more difficulty with a sentence that includes a singular noun phrase subject in the intervening region because it will have a higher activation level than the target noun phrase. For this reason the sentence in Example 2.2 is predicted to be difficult for a cue-based retrieval parser, because *warehouse* now has the syntactic properties sought by the retrieval cue.

Example 2.2: The social worker was surprised that the resident who said that the warehouse was dangerous was complaining about the investigation.

This example constitutes a case of *syntactic interference* where the syntactic properties of the intervening region create retrieval difficulty because the retrieval cues do not unambiguously identify the necessary item. While this effect is predicted by a cue-based retrieval mechanism, it is also predicted by a storage account that includes a limit on the number of similar items that can be stored without penalty (i.e., Lewis (1996)). Consequently, the real test of retrieval effects in these sentences comes in cases with *semantic interference*, when the semantic properties of the intervening region prevent unambiguous identification of the required item, as in sentences Example 2.3.

Example 2.3: The social worker was surprised that the resident who was living near the dangerous neighbor was complaining about the investigation.

This condition is the critical condition for testing both the cue-based retrieval and the global matching hypotheses. A grammatically driven parser that does not incorporate cue-based retrieval would not predict the NP *neighbor* to interfere with the search for a subject for *was complaining* because grammatically it is a prepositional object. Furthermore, a storage account of processing breakdown does not predict the NP *neighbor* to cause any more difficulty than the NP *warehouse* in these sentences, because both place equal demands on storage capacity. Thus, it is only a retrieval account which explains why the semantic properties of *the neighbor* cause it to interfere with the retrieval of the correct subject *the resident.*

In addition, the retrieval account also predicts that syntactic and semantic interference effects will interact because global matching suggests that all cues are equally important and combined multiplicatively into a single retrieval probe. Consequently, the presence of both syntactic and semantic interference results in a much higher probability that the incorrect NP will be retrieved. To illustrate this point, first consider an example SAM retrieval structure patterned after the one presented in Figure 2.1 for a low syntactic and low semantic interference sentence as in Figure 2.3. Here, I have denoted the syntactic and semantic retrieval cues provided by the verb as Q_1 and Q_2 respectively, and self-strength is represented as Q_3. Notice that the self-strength for these NPs increases with their proximity to the retrieval probe (i.e. the final verb), capturing effects of recency/decay.

			Memory Item		
			NP_1	NP_2	NP_3
Cues	Q_1	$\text{Syntax}_{\text{subject, sing, N}}$	.92	.92	.24
	Q_2	$\text{Semantics}_{\text{complain}}$	.84	.89	.19
	Q_3	$\text{Self-strength}_{\text{recency}}$	.60	.70	.80
Probability of Retrieval			.43	.53	.03

Figure 2.3: Retrieval structure for a low syntactic and low semantic interfering sentence. Probability of retrieval calculated with Equation 2.1.

Since the verb is specifically looking for a NP subject, only three items from the sentence are appropriate candidates, *worker, resident,* and *warehouse*, which I have denoted as NP_1, NP_2, and NP_3, respectively. I have estimated the cue-to-image strength for each of these NPs by assessing how well the cues match the NP. For example, the word *warehouse* is the NP_3 and in the syntactic structure of sentence Example 2.1 it appears as a prepositional object and not a subject. Consequently, it will not match the retrieval cue of the verb and is represented here with a low retrieval strength. Similarly, *warehouse* does not fit semantically with the verb *complain*, and is represented with a low retrieval strength. Conjoining these cues multiplicatively according to the formula given in Equation 2.1, suggests that NP_3 will be retrieved with a probability equal to 3%. In fact, the correct noun phrase, NP_2 has the highest probability of retrieval among all NPs at 53% as compared with NP_1 at 43%, which fits with the designation of this sentence as a low interference sentence.[4]

Notice that if either the syntactic or the semantic properties of NP_3 are changed so that their retrieval strength is increased, this will result in increasing the probability of retrieving NP_3 over NP_2 because even if the increase does not surpass the strength value for matching NP_2, the fact that NP_3 is more recent increases the probability of its retrieval. If for example, we increase either the syntactic or the semantic retrieval strength of NP_3 to .90 while keeping all other retrieval strengths at their levels in Figure 2.3, then in both cases this NP_3 now has a higher probability of being retrieved than the 3% probability of NP_3 in the low interference retrieval structure of Figure 2.3. If it is the syntactic interference that is increased to .90, then the probability of retrieving NP_3 becomes 12%, and if it is the semantic interference that is increased, the probability of retrieving NP_3 is 14%

When both are increased, we have the situation represented in Figure 2.4.

			Memory Item		
Cues			NP_1	NP_2	NP_3
	Q_1	$\text{Syntax}_{subject}$	.92	.92	.92
	Q_2	$\text{Semantics}_{complain}$	.84	.89	.90
	Q_3	$\text{Self-strength}_{recency}$	.60	.70	.80
Probability of Retrieval			.28	.34	.39

Figure 2.4: Retrieval structure for a high syntactic and high semantic interfering sentence.

[4] Although the actual retrieval strengths in this example are ad hoc, the precise values are not as important as the mismatch of the NP with the cues, and hence the same result would be predicted as long as this mismatch is preserved. Since syntactic roles are categorical (i.e., and item either is or is not a subject, object, etc.) the syntactic structure defines the extent of the mismatch more precisely than in the case of semantic cues, where the NP could be a better or worse match with the semantics of the retrieval prompt. Further experiments are planned to investigate the parameters of the semantic mismatch effect, including effects of NP-VP frequency and co-occurrence.

In Figure 2.4, all NPs are subjects and hence have the same retrieval strength vis-à-vis the subject retrieval cue. The semantics of NP_3 also fits with the verb *complain*, meaning that there is both syntactic and semantic interference in this sentence. When the retrieval formula is applied to this example, we find that the probability of retrieving NP_3, the incorrect NP, is 39%, which is higher than retrieving the correct NP_2, which has a retrieval probability of only 34%. Thus, the probability of retrieving an intermediate, and interfering NP increases linearly as the amount of interference vis-à-vis the retrieval cues of the probe item increases. We expect to see this effect empirically by observing a significant interaction of syntactic and semantic interference, as well as effects of both syntactic and semantic interference in the experiments conducted in the remainder of this chapter.

2.4 Materials

In order to test the cue-based retrieval and global matching hypotheses discussed above, one hundred sixty sets of items were constructed, each with 5 forms exemplifying 5 experimental conditions. An example is given in Table 2.1 below, with the intervening regions for each of the five conditions presented in sequence. The five conditions can be constructed by combining the sentence introduction and conclusion with one of the 5 intervening regions.

Table 2.1: Example syntactic and semantic interference stimuli.

Sentence introduction	The social worker was surprised that
Intervening region: short	the resident
Intervening region: low syntactic interference/low semantic interference	the resident who was living near the dangerous warehouse
Intervening region: low syntactic interference/high semantic interference	the resident who was living near the dangerous neighbor
Intervening region: high syntactic interference/low semantic interference	the resident who said that the warehouse was dangerous
Intervening region: high syntactic interference/high semantic interference	the resident who said that the neighbor was dangerous
Sentence conclusion	was complaining about the investigation.

The short region provides a control condition for making the syntactic attachment without effects due to distance or decay. The syntactic and semantic interference conditions were constructed according to the theoretical predictions described above.

2.4.1 Syntactic interference content norming

In order to determine whether any observed differences between the high and low syntactic interference conditions could be due simply to the different content in the intervening region, a content norming experiment was conducted in which the syntactic structure of the sentences was altered, while keeping the content identical to that found in the experimental sentences.

2.4.1.1 Subjects

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

2.4.1.2 Materials

A randomly selected set of 36 experimental items from the above described set was selected. Content match sentences were constructed by transforming the interfering sentences into the content match sentences shown in Table 2.2.

Table 2.2: Content match sentences for syntactic interference pilot.

Syntactic interference condition	Interfering sentence	Content match
Low interference	The social worker was surprised that the resident who was living near the dangerous warehouse was complaining about the investigation.	The resident was living near the dangerous warehouse, and the social worker was surprised that she was complaining about the investigation.
High interference	The social worker was surprised that the resident who said that the warehouse was dangerous was complaining about the investigation.	The resident said that the warehouse was dangerous, and the social worker was surprised that she was complaining about the investigation.

2.4.1.3 Procedure

The 36 content matched sentences were presented in a pseudo-random order wherein each experimental item was blocked between 2 filler items. Filler sentences contained a set of 2 clause sentences separated by a comma. The experiment was run on Pentium-class personal computers using the MEL2 Professional experimental package (Schneider, 1995). Sentences were presented in a non-cumulative, self-paced, moving window format, where each sentence was presented one word at a time. Prior to the experiment, subjects were presented with the following instructions:

> In this experiment you are asked to read a series of English sentences and to decide whether or not they are grammatical. A grammatical sentence is a sentence that satisfies the rules of English grammar (syntax). For example, the sentence *the man likes football* is a grammatical sentence but we

would not say *the man likes* is grammatical because there is something missing from the sentence. Similarly *the man waits the lady* is ungrammatical because *the lady* is extra in the sentence.

When they were ready to begin, subjects pressed the space bar and 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 the 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.

2.4.1.4 Results and Discussion

Accuracy scores for the low interference content matches were 79% and 73% for high interference content matches. While judgments appear quite low for these sentences, this is likely 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 processing differences.

2.4.2 Semantic interference pilot

The semantic interference conditions also required piloting because it was necessary to ensure that the interfering noun could serve as a plausible subject for the final verb phrase. In order to verify that these sentences were semantically interfering, we conducted a pilot experiment in which these sentences were transformed into the three conditions presented in Table 2.3.

Table 2.3: Sentences for semantic interference pilot

Condition	Sentence
Short	The social worker was surprised that the resident was complaining about the investigation.
Implausible	The social worker was surprised that the meeting was complaining about the investigation.
Plausible	The social worker was surprised that the neighbor was complaining about the investigation.

These sentences provide an appropriate test of semantic interference in the experimental items in Table 2.1 because semantic interference only occurs when the cues provided by the final verb match the features of the intervening NP sufficiently well that they could be construed as the subject of that verb. Consequently, if the Plausible and Short conditions in Table 2.3 are rated as similarly plausible, then we would expect to observe semantic interference when these NP-Verb combinations appear in the structures in Table 2.1. Similarly, for semantic interference to be absent, the implausible condition must be recognized as such, and be significantly different from either the plausible or the short conditions.

2.4.2.1 Subjects

Subjects from the University of Pittsburgh undergraduate psychology subject pool participated in the experiment for partial credit. Sixteen subjects rated the first half of the stimuli and fifteen subjects rated the second (see Procedure below).

2.4.2.2 Procedure

The three conditions of the 160 item sets produced 480 sentences that required piloting. I split this list into two so that any one subject would be required to rate only 240 sentences in a one-hour testing period. The sentences were presented in a random order differing for each subject.

Prior to the experiment, subjects were presented with the following instructions:

In this experiment you will be asked to read a series of sentences and rate each one for how well it makes sense. You will use a 5-point scale, where 5 means it makes perfect sense and 1 means it makes no sense at all. For example, imagine you are given the following sentence:

The lamp was walking slowly down the street.

(NONSENSE) 1 - 2 - 3 - 4 - 5 (MAKES SENSE)

Since lamps don't usually walk down the street, you would probably rate this sentence with a '1'. On the other hand, if the sentence were:

The disappointed boy was walking slowly down the street.

(NONSENSE) 1 - 2 - 3 - 4 - 5 (MAKES SENSE)

You would probably rate this sentence with a '5'.

During the experiment, each sentence appeared on the screen in its entirety, followed by the same rating scale seen above in the instructions. Subjects entered their rating for that sentence and then pressed the space bar to move to the next sentence. Subjects had no time pressure and received no feedback in the experiment, and no filler items were included.

2.4.3 Results

Two One-Way ANOVAs were conducted on each item separately. The first compared the two plausible conditions, and we expected to find no significant difference between these two. Fifty-two of the 160 items did produce a significant difference between the two plausible conditions, and therefore required revision before they could be used to test semantic interference effects. The second ANOVA compared the plausible and the implausible conditions, and we did expect to find a significant difference between them. Eighteen of the 160 item sets did not have a statistically significant difference between the plausible and implausible versions, indicating that these sentences also required revision. Thirteen of these 18 also did not fit the prediction of no significant difference between the two plausible conditions.

The problematic sentences were corrected in one of three ways. First, items that could be corrected in an obvious way, for example, choosing a different NP that would be implausible as the subject of the verb, were altered. Second, a new set of 16 items was constructed and piloted with 12 subjects. This pilot was conducted more informally, using subjects associated with the LRDC and was conducted on paper. Each subject rated each of the 48 items using the same scale used in the original pilot experiment and the data were analyzed in the same way as described here. Those in this set which satisfied the appropriate criteria, or which could be corrected in an obvious way, were used as replacements for the problematic sentences in the original set. Lastly, 10 items were thrown out entirely, leaving a pool of 150 experimental items from which the materials in Experiments 1 and 2 were drawn.

2.5 Experiment 1: Got It? Task

A preliminary test of the syntactic and semantic interference effects was conducted using the Got It? task (Frazier, Clifton, & Randall, 1983). This task was chosen because it provides an indication of the interpretability of sentences without requiring subjects to make explicit decisions about grammar, a topic that appears to cause anxiety in otherwise capable students. In this task, subjects are instructed to judge as quickly as possible after the end of the sentence whether or not they understood the sentence. If the subject had no difficulty understanding the sentence, they will answer 'Yes'. A 'No' answer indicates that the subject was unable to make sense of the sentence, and in the case of our interfering sentences indicates that subjects had difficulty identifying the proper subject for the final verb so that they could complete the long distance attachment. It is predicted therefore that both syntactic and semantic interference will have a detrimental effect on got it scores, as will their combination. In addition, if the distance over which the long distance attachment must be made affects the ability to identify the subject, then a sentence with low syntactic and low semantic interference in the intervening region should nevertheless be more difficult to understand than the short sentence in which the attachment can be made immediately. Thus, using the materials in Table 2.1 the effects due to syntactic interference, semantic interference, and distance can be tested simultaneously.

2.5.1 Method

2.5.1.1 Subjects

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

2.5.1.2 Materials

Fifty sets of experimental items were chosen from the pool described in Section 2.4 for use in this experiment. Five lists of items were constructed so that each subject received one of the 5 conditions from each set of items, but no subject saw all the conditions for any one set. Each subject received sentences in each of the 5 conditions, permitting a within-subjects analysis of the data. Items were presented in a blocked random order such that every experimental item was separated by three filler items of different syntactic constructions. Thus, a total of 200 sentences were presented to each subject.

The filler items were designed to be appropriate matches for several aspects of the experimental sentences. In order to match the structures of our interfering items, in half of the filler items we used subject relative clauses as objects, (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 low interference items except that there was no long distance attachment required. Of the other half, two-thirds were simple 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 focusing on length as indicating a hard to comprehend sentence.

2.5.1.3 Procedure

As with the pilot experiment, this experiment was implemented in the MEL2 Professional experimental package (Schneider, 1995) and was run on Pentium-class personal computers. Excluding the instructional content, the same procedure as described in Section 2.4.1.3 was used here. Prior to the experiment, subjects were presented with the following instructions:

> In this experiment you are asked to read a series of English sentences and indicate whether or not you have understood the sentence by answering Yes or No to the question 'Did you get it?' You should answer this question as quickly as you can, without trying to make sense of a sentence that sounds awkward. Some sentences in this experiment are designed to be difficult to understand.

2.5.1.4 Design and analysis

The experiment was analyzed as a 2 (high/low syntactic interference) x 2 (high/low semantic interference) factorial repeated-measures ANOVA. In addition, paired t-tests were conducted comparing each interference condition with the short condition to test the effect due to simply making the long distance attachment.

2.5.2 Results

Table 2.4 presents the percentage of the sentences in each condition for which subjects said that they did "Get it," indicating that they had no trouble understanding the sentence.

Table 2.4: Accuracy Scores (standard error) in Got It? Task, Experiment 2.3

	Syntactic Interference	
Semantic Interference	NO	YES
NO	.91 (SE_1 = .02; SE_2 = .02)	.83 (SE_1 = .03; SE_2 = .03)
YES	.80 (SE_1 = .03; SE_2 = .03)	.78 (SE_1 = .03; SE_2 = .02)
Short Condition	.95 (SE_1 = .01; SE_2 = .01)	

A main effect of syntactic interference was found, F_1 (1,34) = 18.85, p < .001; F_2 (1,49) = 10.17, p < .005, with the high interfering sentences being more difficult than the low interference sentences. Similarly, a main effect of semantic interference was found, F_1 (1,34) = 9.84, p < .005; F_2 (1,49) = 5.87, p < .02. There was a slight suggestion that the interaction might be significant by subjects, F_1(1,34) = 2.91, p < .10, however, this was not supported by the items analysis, F_2 (1,49) = 1.24, p < .27.

Pairwise comparisons with the short condition were all significantly different in the subject analyses, as illustrated in Table 2.5. The short condition compared with the low syntactic/low semantic interference condition was not significant in the items analysis.

Table 2.5: Pairwise comparisons of Interference conditions and short condition.

Short condition compared with:	t-test results
Low syntactic/Low semantic interference	t_1(34) = 2.16, p < .04; t_2(34) = 1.79, p < .08
Low syntactic/High semantic interference	t_1(34) = 4.44, p < .001; t_2(34) = 4.52, p < .001
High syntactic/Low semantic interference	t_1(34) = 4.85, p < .001; t_2(34) = 5.41, p < .001
High syntactic/High semantic interference	t_1(34) = 6.24, p < .001; t_2(34) = 6.90, p < .001

2.5.3 Discussion

Despite the fact that the "Got it?" task provides only an off-line test of the effects being explored here, we did find evidence for both the syntactic and the semantic interference effects, as well as a distance effect in this experiment. These effects support the prediction that properties affecting retrieval make a sentence more difficult to process. This is particularly the case with the semantic interference effect, which is not predicted by storage accounts of processing limitations. Moreover, a purely structural account of these complexity effects, arguing that limitations come from the center-embedded clauses, or incomplete syntactic predictions cannot account for this data, since semantic effects were observed regardless of the whether the syntactic structure was easy or difficult to process.

The distance effect observed here supports an account in which items become more difficult to retrieve as they become more separated from the retrieval probe. This could be explained either by interference from the intervening material or else simply by decay of less recent items. The data observed here provide some support for the decay account over the interference account since the distance effect was observed even when the intervening material did not match the properties of the retrieval probe (i.e., in the low syntactic and low semantic interfering conditions). However, this effect was significant only by subjects, so an interference account can not be completely dismissed. This suggests that retrieval of the target item is not influenced solely by self-strength, and may be influenced by even partial matches with intervening items. This is predicted by the principle of global matching, since the retrieval cues are compared with all items in working memory, and not simply the "best matches" for the cues (cf. the denominator in Equation 2.1, which sums over all memory items).

One aspect of these data that must be explained further is the non-significant interaction between semantic and syntactic interference, and the hint that the semantic interference effect may disappear in the context of syntactic interference. This result is contrary to the prediction of the SAM memory model regarding how cues combine to create retrieval prompts, as it implies that cues may not be combined with equal weight in the retrieval probe. I suggest that this may be due to the off-line and loosely constrained nature of the Got it? task, and will return to this issue in Experiment 2.

2.6 Experiment 2: Reading Comprehension Task

The Got It? task from Experiment 1 provided preliminary evidence for the syntactic and semantic interference effects, as well as effects of decay. A major drawback with this experiment, however, is the fact that the task did not provide an online assessment of these effects. The current experiment attempts to address this by using a word-by-word reading task, producing reading times in four designated regions of interest. Notably, this will also permit a direct comparison between the retrieval interference account of complexity effects, and a storage account of these effects, such as the Discourse Locality Theory (DLT) (Gibson, 2000; Warren & Gibson, in press). Warren and Gibson (in press) have argued that complexity effects can be explained by a metric based on 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) grammatical heads and dependents; and 3) those associated with storing incomplete structural dependencies (i.e. syntactic predictions). These costs are illustrated in Table 2.6 for the low syntactic/ low semantic interference sentence in Example 2.4, and in Table 2.7 for the high syntactic/low semantic interference sentence in Example 2.5.

Example 2.4: The older boy understood that the girl who was swimming near the dock was paranoid about dying.

Example 2.5: The older boy understood that the girl who said that the townspeople were dangerous was paranoid about dying.

Table 2.6: Word by word predictions of the Discourse Locality Theory for a low syntactic interfering sentence.

	Input word															
Cost type	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	0	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 2.7: Word by word predictions of the Discourse Locality Theory for a high syntactic interfering sentence.

	Input word															
Cost type	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

The DLT assigns 1 energy unit (EU) for each new contentful discourse referent, which in the case of Example 2.4 includes nouns, verbs, and prepositions.[5] This is illustrated in the first row of Table 2.6. 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 2.6. 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, however, 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 decremented when the NP *dock* satisfies the prepositional prediction, leaving only the outstanding verbal prediction, which is finally realized when *paranoid* occurs.

[5] 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.

In like fashion, Table 2.7 presents the DLT costs for the high interference sentence in Example 2.5. 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 Example 2.5 are more difficult than low interfering sentences like Example 2.4. The shaded regions in Table 2.6 and Table 2.7 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 Example 2.4 and Example 2.5 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. Thus, reading times will allow distinguishing a retrieval theory such as the Interference theory from a storage theory such as the DLT. This is important because the difference between the high and low interfering sentences could also be due to *storage interference*, as argued by Lewis (1996) (cf. Section 1.2.4). He suggested that items occurring in similar syntactic positions are stored with a syntax-specific encoding, and that interference arising from the storage of multiple similar items produces processing breakdown. On this account, the complexity difference between Example 2.4 and Example 2.5 is due to the storage of the three verb phrases *understood*, *said*, and *were* in the high interference sentence, but only two in the low interference sentence, *understood,* and *was (swimming).* Lewis marshals other data suggesting a limit of two (or three) on verbal short-term memory (e.g., (Daneman & Carpenter, 1980; Gibson, 1991; Simon & Zhang, 1985)) and concludes that by the time the fourth VP *was (paranoid)* occurs, the parser can no longer access the *understood* VP which is required for completing the attachment of the complement clause headed by *was (paranoid)*. As with the DLT, if storage costs contribute to processing breakdown, then evidence of their effect should be observable at the point of highest storage complexity, i.e. at the verb *were (dangerous)* in the high interfering sentences. Note that this is the same region that the DLT predicts will exhibit processing differences, and notably, it is earlier than the region predicted to cause difficulty according to the retrieval interference account described here.

2.6.1 Method

2.6.1.1 Subjects

Forty-eight 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.

2.6.1.2 Materials

Thirty-two items from the pool of materials described in Section 2.2 were used here affording a 2 x 2 comparison of syntactic and semantic interference. In addition, a distance manipulation was added so that the effect of decay on the interfering NP can be tested. The difference between the short and the long sentences in Experiment 1 indicated that recency does affect the easy with which items can be retrieved in order to make the correct attachment. In addition, it could be argued that the semantic interference effect is due to the semantically plausible NP in the intervening region being adjacent to the final verb (i.e. recent), and not due to the effects of cue-based retrieval. To evaluate this explanation, we created a distance manipulation that would preserve the property of semantic interference while moving the interfering NP further away from the verb prompting the retrieval. This involved adding a prepositional phrase after the intervening NP in all conditions. Table 2.8 shows an example of all 8 conditions.

Table 2.8: Materials for Experiment 2.

Interference condition	No distance manipulation (Materials from Experiment 1)	With Distance Manipulation
Low syntactic, low semantic	The // worker was surprised that the resident who / was living near the dangerous warehouse / had complained / about the // investigation.	The // worker was surprised that the resident who / was living near the dangerous warehouse on the corner / had complained / about the / investigation.
Low syntactic, high semantic	The // worker was surprised that the resident who / was living near the dangerous neighbor / had complained / about the // investigation.	The // worker was surprised that the resident who / was living near the dangerous neighbor on the corner / had complained / about the // investigation.
High syntactic, low semantic	The // worker was surprised that the resident who / said that the warehouse was dangerous / had complained / about the / investigation.	The // worker was surprised that the resident who / said that the warehouse on the corner was dangerous / had complained / about the // investigation.
High syntactic, high semantic	The // worker was surprised that the resident who / said that the neighbor was dangerous / had complained / about the // investigation.	The // worker was surprised that the resident who / said that the neighbor on the corner was dangerous / had complained / about the // investigation.

In all cases the distance manipulation was designed so as not to add any additional interference into the intervening region. Syntactically, the NP is not a subject NP and therefore will not match the syntactic retrieval cues of the final verb. Semantically, the NP was chosen so as to be implausible as the subject of the final verb, which in most cases required choosing an inanimate NP, but we also relied on the a notion of "semantic fit" like that used in creating or semantic interference sentences.

2.6.1.3 Procedure

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. Subjects responded using “yes” and “no” keys on the keyboard, designated as the “1” and “2” keys.

Two types of comprehension questions were used for the experimental sentences. The first pertained to the long distance attachment required for integrating the final verb into the sentence, and therefore prompted a "Yes" response. For example, the comprehension question for the example in Table 2.8 would be "*Was it the resident who complained?*" We will refer to these questions as target-NP questions. In order to prevent any learning that could occur from seeing the target-NP questions that have the correct attachment preserved, we used non-target-NP questions for the second half of the experimental stimuli. These sentences queried one of the other NPs in the sentence and always required a "No" response. In low semantic interference sentences, this entailed substituting the matrix subject NP as the subject of the final verb, e.g. *"Was it the worker who complained?"* In the high semantic interference sentences the interfering NP was substituted, e.g., *"Was it the neighbor who complained?"* This permitted an evaluation of the extent to which any prior subject NP contributes to retrieval interference, or whether it is mainly the subject in the region intervening for the long distance attachment that makes the sentence difficult.

2.6.1.4 Dependent measures & Analysis

This experiment affords two separate dependent measures for evaluating the effects of interest. Accuracy scores to comprehension questions provide an off-line measure of processing difficulty similar to the Got it? task in Experiment 1, but with a task that required subjects to pay much closer attention to the meaning of the sentences, since they knew their comprehension would be tested with a question specific to the sentence they had just read. The target-NP and non-target-NP questions were analyzed separately, using a 2 (syntactic interference) x 2 (semantic interference) x 2 (distance) within subjects ANOVA on accuracy scores. Due to an error in materials preparation, the two types of comprehension questions were not evenly distributed across the 8 conditions. This made a repeated measures ANOVA with subjects as the random variable impossible, so this analysis was conducted only by items.

The second dependent measure was reading time per region during the self-paced reading task. Sentences were partitioned into four regions of interest, as illustrated by slashes in the sentences in Table

2.8. The first and the last words of each sentence, separated from the other regions by double slashes in Table 2.8, were not analyzed because the reading times for these words are assumed to include artifactual slowdowns, such as preparatory effects and sentence-final wrap up effects (cf. (Just & Carpenter, 1980; King & Just, 1991) for similar technique).

The first region of interest contained the subject NP, the matrix verb, and the initial section of the embedded relative clause. The second region denoted the intervening region, including the distance manipulation. Note that this is the region in which we expect to find evidence of any effect due to additional storage demands for the high syntactic interfering sentences. Region three was the critical region, as it contained the VP for the long distance attachment and is therefore predicted to be the locus of retrieval interference effects. Region four was a final spillover region, containing up to two words after the critical region, which in most cases constituted a prepositional phrase.

Analyses of the data for each region were conducted on residual reading times, which enabled removal of variance associated with subject-specific reading rates and differences in region lengths. A regression equation for each subject was calculated using word length as a predictor 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 to costs associated with the button-press in a self-paced reading paradigm. The constant in 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, a word position predictor was included in the regression analyses to correct for a gradual increase 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 (e.g., (Ferreira & Henderson, 1993; Gernsbacher & Faust, 1990; Sturt, Pickering, Scheepers, & Crocker, 2001)) none have attempted to statistically correct for the effect (but see Zwaan, Graesser, and Magliano (1995) who report regression analyses on whole-sentence reading times suggesting that subjects do read faster as they progress through a text.) In this analysis, such a correction is essential since I will be comparing a critical region occurring as the 14-16th words in sentences with no distance manipulation and as the 18-20th words in the longer sentences. Moreover, the regression equations for individual subjects reveal that 38 of the 48 subjects had negative coefficients for the word position predictor (M = -4.21), suggesting that these subjects' reading time did increase as they moved through the sentence. Table 2.9 presents the average coefficients for the predictors in each equation, as well as t-tests indicating that they are all significantly different from zero.

Table 2.9: Coefficients for regression predictors.

Predictor	Mean	SD	t-test	p-value
Constant	366.61	85.52	t(47) = 29.70	p < .001
Position	-4.21	4.66	t(47) = -4.21	p < .001
Word length	16.03	9.63	t(47 = 11.54	p < .001

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 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, Pickering, Scheepers, and Crocker (2001) we trimmed outliers in each region by replacing the values that were either 3 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 less than 2% of the data.

2.6.2 Comprehension Question Results & Discussion

2.6.2.1 Yes (Target-NP) questions

Table 2.10 presents accuracy scores for the Target-NP questions. Main effects of both syntactic and semantic interference were present with these questions, $F_2(1,17) = 27.95$, $p < .001$ for syntactic interference and $F_2(1,17) = 18.24$, $p < .002$ for semantic interference. The interaction between syntactic and semantic interference was significant, $F_2(1,17) = 6.01$, $p < .03$, as predicted by the global-matching hypothesis.

The main effect of distance was not significant, however, there was a significant interaction between distance and semantic interference, $F_2(1,17) = 4.97$, $p < .05$. This interaction is due to the distance manipulation having no effect in the presence of low semantic interference, where accuracy scores were 85% regardless of the distance manipulation. In contrast, when semantic interference was present, accuracy was 73% without the distance manipulation and 65% with the manipulation. The lack of a distance effect when there is no semantic interference reflects the fact that there is little cause for incorrect interpretation arising from the retrieval of the incorrect NP, and therefore a subject's ability to answer the question is not hindered. Yet when semantic interference is present, subjects are more likely to have created an incorrect interpretation of the sentence. Their error becomes evident when they are presented with the target question, which effectively supplies the correct interpretation. In such a case, subjects' must retrieve the proper NP to check its predicate, and this is more difficult if the NP has decayed more because of the distance manipulation. Similarly, there was a hint of the same sort of interaction between distance and syntactic interference, $F_2(1,17) = 3.19$, $p < .10$. As above, this interaction was due to a non-significant effect of distance in the presence of low syntactic interference, with accuracy scores at 83% without the distance manipulation and 84% with the distance manipulation for low syntactic interference sentences. Without syntactic interference, the probability of an incorrect interpretation is lower so subjects can answer the comprehension question at the same rate as when there was no semantic interference. When syntactic interference is present, however, subjects again are forced to retrieve the correct NP to check its predicate. When the distance manipulation is not present, this is easier, as seen with accuracy scores of 74%, but with the distance manipulation this becomes much more difficult, yielding accuracy scores of 66%.

Table 2.10: Accuracy for Target-NP questions in Experiment 2

		Accuracy
No distance manipulation Interference Type		
	Low syntactic, Low semantic	.88 (.03)
	Low syntactic, High semantic	.79 (.04)
	High syntactic, Low semantic	.82 (.04)
	High syntactic, High semantic	.67 (.05)
With distance manipulation Interference Type		
	Low syntactic, Low semantic	.89 (.03)
	Low syntactic, High semantic	.79 (.05)
	High syntactic, Low semantic	.81 (.05)
	High syntactic, High semantic	.51 (.05)

2.6.2.2 No (Non-target-NP) Questions

Recall that the form of this question differed depending on whether semantic interference was present or not. If there was semantic interference, then the interfering NP was paired with the final verb providing the retrieval cue. If there was low semantic interference, then the matrix NP was presented as the subject of this verb. If these questions were ruled out at a similar rate, we can conclude that it is merely the presence of other NPs in the sentence that make it confusing, and not the relative position of these NPs *vis-à-vis* the retrieval probe. This was found to be not the case in our data (see Table 2.11), as a main effect of question type was found, $F_2(1,13) = 16.14$, $p < .002$. When the semantically interfering NP was the subject, the question was much more difficult to reject as evident by a 55% accuracy value, meaning that subjects often believed that the interfering NP *was* the proper subject of the final verb. When the matrix NP was the subject, the question could be rejected at the rate of 76%, suggesting that the relative position of these NPs may play a role in whether they are more likely to be interpreted as the subject of the verbal retrieval probe. No other effects were significant in this analysis.

Table 2.11: Accuracy for Non-target-NP Questions

		Accuracy
No distance manipulation Interference Type		
	Low syntactic, Low semantic	.80 (.04)
	Low syntactic, High semantic	.60 (.06)
	High syntactic, Low semantic	.79 (.05)
	High syntactic, High semantic	.57 (.08)
With distance manipulation Interference Type		
	Low syntactic, Low semantic	.76 (.05)
	Low syntactic, High semantic	.57 (.08)
	High syntactic, Low semantic	.68 (.08)
	High syntactic, High semantic	.45 (.05)

2.6.3 Reading times Results & Discussion

The comprehension question data above shows that subjects had an overall accuracy of 77% for target-NP questions and 65% for non-target-NP questions, suggesting that their ability to construct accurate interpretations for these sentences was rather low. Because of this, I analyzed only those items for which subjects correctly answered their comprehension question. As a result of missing data because a given subject got all sentences in a particular condition incorrect, 4 subjects were discarded from the analyses. Similarly, in one of the experimental items, one of the conditions was always answered incorrectly, so this item was discarded. These procedures resulted in different means for the subject and item analyses, presented in Table 2.12 and Table 2.13, respectively. Table A.1 and Table A.2 in Appendix A present raw reading times (i.e., without regression analyses) for the regions of interest, trimmed according to the same procedures discussed above. This trimming affected less than 2% of the raw data. Statistical analyses on these raw reading times produced results parallel to those reported here, as summarized in the Appendix.

No significant effects were found in Region 1. In Region 2, no significant effects were found in the subjects analyses, but a significant effect of syntactic interference was found in the items analysis,

$F_2(1,30) = 4.35$, $p < .05$. This is due to longer reading times in the high syntactic interference condition (M=527.73, SE = 126.81) than when syntactic interference was reduced (M = 372.66, SE = 89.20). Although weak, especially since it is not at all indicated in the subjects analysis, $F_1(1,43) = .94$, $p < .34$, this result fits with the prediction of Gibson's DLT model. However, it also fits with the retrieval interference model since the high syntactic interference sentences require an intermediate retrieval in this region in order to complete the embedded clause. Consequently, this result alone cannot distinguish the two theories.

In region 3, a syntactic interference effect was found for both subject and items analyses, $F_1(1,43) = 7.65$, $p < .009$; $F_2(1,30) = 8.29$, $p < .008$. As expected, this effect is due to greater difficulties when the syntactic interference is high ($M_1 = 149.47$, $M_2 = 152.75$) than when it is low ($M_1 = 87.35$, $M_2 = 75.29$). A significant main effect of semantic interference was also found, $F_1(1,43) = 12.72$, $p < .002$; $F_2(1,30) = 10.01$, $p < .005$. Again, this was due to conditions where semantic interference was high (M_1=154.40, M_2 = 152.19) being more difficult than when semantic interference was low ($M_1 = 82.42$, $M_2 = 75.85$). The syntactic x semantic interference interaction showed a trend for significance in the subjects analysis, $F_1(1,43) = 3.75$, $p = .06$, and did reach significance in the items analysis, $F_2(1,30) = 5.69$, $p < .03$.

The distance manipulation in Region 3 was not significant in either the subject or items analysis, Fs < 1. This effect suggests that the interference effects observed previously are not due simply to the proximity of the distracting NP to the retrieval cue. Although the low syntactic interference conditions have the distracting NP positions adjacent to the retrieval cue when there is no distance manipulation, this is not the case in the conditions with the distance manipulation, so it is unlikely that subjects are using an "adjacent NP" strategy instead of conducting a cue-based retrieval to find the proper subject for this verb.

Region 4 consisted of the "spill-over" region occurring after the location of the subject retrieval. In this region, a significant effect of syntactic interference was found, $F_1(1,43) = 8.06$, $p < .008$; $F_2(1,30) = 8.96$, $p < .006$. Likewise, a significant effect of distance was found in this region, $F_1(1,43) = 19.23$, $p <$

.001; $F_2(1,30) = 31.31$, $p < .001$. No other effects, including the semantic interference effect, were significant in this region. In order to understand these two effects, it must be noted that they are not related to the retrieval that must occur in Region 3, as there is no reason to believe that retrieval effects should surface in Region 4, especially since effects on retrieval such as the semantic interference effect have been observed at the retrieval point itself. Rather these effects must be due to the sentence processing operations occurring after the long distance attachment has occurred, namely those processes related to the creation of a situational or referential model that captures the subjects global interpretation of the sentence's meaning (Garrod & Terras, 2000; Haberlandt & Graesser, 1989; Zwaan et al., 1995). The question to ask about these results then, is why syntactic interference and distance should effect the processes related to creating this model. This distance effect can be explained by noting that with the distance manipulation reading times are slower in this region ($M_1 = 109.46$, $M_2 = 108.86$) than when there is no distance manipulation ($M_1 = 55.98$, $M_2 = 45.42$). This is as would be expected if the process of creating a situational-referential model requires accessing the content of previously occurring sentence items, since the distance manipulation necessitates that all previous items have had more time to decay than when there was no such manipulation. In fact, the syntactic interference effect observed here can also be understood in this light, since the high interfering syntactic structure has the intermediate NP occurring 5 words prior to this point, while the low interfering syntactic structures have this intermediate NP occurring only 3 words prior. Hence the NP referent in the high interference condition has had more time to decay, resulting in more difficulty accessing its content and higher reading times ($M_1 = 101.25$; $M_2 = 94.24$) than when the NP has had less time to decay ($M_1 = 64.19$; $M_2 = 60.04$).

Table 2.12: Trimmed residual reading times for each region (ms), subjects as random factor.

		Region 1	Region 2	Region 3	Region 4
No distance manipulation Interference Type					
	Low syntactic, Low semantic	-64.18 (56.00)	327.47 (66.39)	86.39 (22.85)	43.47 (15.70)
	Low syntactic, High semantic	-89.48 (62.91)	291.54 (98.15)	112.59 (30.18)	38.53 (15.61)
	High syntactic, Low semantic	-39.46 (52.62)	512.77 (99.91)	112.46 (29.86)	59.48 (19.43)
	High syntactic, High semantic	-34.03 (79.91)	506.08 (140.20)	197.86 (35.06)	82.45 (20.17
With distance manipulation Interference Type					
	Low syntactic, Low semantic	-33.62 (54.31)	499.90 (126.38)	53.42 (25.15)	91.51 (15.60)
	Low syntactic, High semantic	-124. 06 (59.41)	516.25 (140.47)	96.99 (30.58)	83.26 (17.33)
	High syntactic, Low semantic	-109.55 (50.93)	438.57 (114.87)	77.42 (27.92)	114.42 (20.14)
	High syntactic, High semantic	-137.81 (67.94)	508.56 (151.91)	210.15 (48.65)	148.66 (28.43)

Table 2.13: Trimmed residual reading times for each region (ms), items as random factor.

		Region 1	Region 2	Region 3	Region 4
No distance manipulation Interference Type					
	Low syntactic, Low semantic	-54.93 (59.12)	323.03 (100.49)	79.52 (33.05)	39.36 (18.71)
	Low syntactic, High semantic	-68.13 (76.68)	267.64 (108.02)	86.78 (31.63)	24.97 (17.44)
	High syntactic, Low semantic	-54.19 (73.45)	479.28 (108.27)	88.46 (35.27)	52.07 (15.15)
	High syntactic, High semantic	-73.07 (68.52)	483.90 (139.12)	195.38 (49.43)	65.29 (22.77)
With distance manipulation Interference Type					
	Low syntactic, Low semantic	-36.37 (67.33)	499.03 (144.58)	56.91 (27.86)	91.74 (17.53)
	Low syntactic, High semantic	-78.10 (71.55)	400.94 (118.58)	77.95 (27.66)	84.09 (18.27)
	High syntactic, Low semantic	-131.76 (57.60)	450.16 (179.60)	78.52 (26.25)	113.35 (20.93)
	High syntactic, High semantic	-110.11 (82.76)	697.58 (216.82)	248.65 (65.33)	146.27 (28.14)

2.7 Conclusions

The experiments conducted here were designed to test the hypothesis that parsing operates according to the memory principles of cue-based retrieval and global matching. The materials used to test this hypothesis allowed for an intervening NP to occur between a long distance subject-verb dependency. It was hypothesized that when the verb in this dependency occurs, it supplies retrieval cues that are used to identify its proper subject from the previously occurring NPs. Consequently, the effect that the intervening NPs have on subjects' ability to complete this long distance attachment provides important data which any proposed parser must explain.

The present data clearly illustrate that an intervening NP with the same grammatical function as that desired by the verb supplying the retrieval cues significantly degrades the ability to complete the long distance dependency. This effect was observed only at the point of cue-based retrieval, and not in the region where the NP occurs, suggesting that the underlying cause of the diminished comprehension is not due to a limit on storage capacity. Consequently, the data presented here argue against storage-based accounts of these complexity effects, including the accounts of Gibson (1998; 2000; Warren & Gibson, in press) and Lewis (1996). In addition, a rule-driven parser operating in a top-down fashion, such as that of Frazier (1987; Frazier & Clifton, 1996) does not predict an effect of the intervening NP on the long distance dependency that must be created after it has already been parsed because this NP has already received its grammatical role. Rather, if items are indexed in a retrieval structure according to their linguistic features, then all previously occurring items may still affect processing at any retrieval point, precisely *because* they have already been parsed and been assigned a grammatical role in the sentence. To illustrate, Figure 2.5 provides a possible retrieval structure for the sentence in Example 2.1 repeated here as Example 2.6, with ad hoc retrieval strength values chosen to reflect how well the linguistic item might match to each given cue.

Example 2.6: The social worker was surprised that the resident who was living near the dangerous warehouse was complaining about the investigation.

Figure 2.5: Retrieval structure for NPs

			Memory Item						
			NP_1	V1	NP2	V2	P	NP3	V4
Indexed Cues	Q_1	$Syntax_{sNoun}$	.99	.01	.95	.01	.01	.98	.01
	Q_2	$Syntax_{Verb}$	.01	.67	.01	.98	.01	.01	.97
	Q_3	$Syntax_{Prep}$	.01	.01	.01	.01	.99	.01	.01
	Q_4	$Syntax_{Subject}$	.92	.03	.92	.02	.02	.02	.02
	Q_5	$Syntax_{V\text{-}object}$	.02	.02	.23	.82	.01	.02	.01
	Q_6	$Syntax_{Prep\text{-}object}$	.01	.02	.03	.01	.02	.97	.03
	Q_7	$Semantics_{surprise}$	.78	.99	.82	.02	.01	.06	.02
	Q_8	$Semantics_{living}$	.91	.04	.07	.99	.45	.04	.02
	Q_9	$Semantics_{near}$	.04	.02	.12	.67	.98	.02	.01
	Q_{10}	$Semantics_{complain}$	.96	.01	.95	.02	.02	.02	.99
	Q_n	Other cues	.76	.34	.13	.02	.11	.70	.65
	Q_{n+1}	$Self\text{-}strength_{recency}$	.30	.40	.50	.60	.70	.80	.90

While the specific values in Figure 2.5 are not important, the figure does illustrate how each memory item has associated with it an indexed feature structure which supports cue-based retrieval. When the head of a grammatical dependency occurs, that head specifies the semantic and grammatical features required for its dependent and these features are combined multiplicatively into a single memory probe. This probe is then compared with a retrieval structure such as that above using the formula given in Equation 2.1, resulting in retrieval of the best matching item and the completion of the grammatical dependency. The success of a parse, and the resultant comprehension of a sentence therefore depends not only on the grammatical structure of the sentence, but on the linguistic properties of all other words in the sentence and how they combine in the retrieval structure.

Further specification of this is provided in the principle of global-matching, which gives a precise definition for how retrieval cues are combined. Namely, each retrieval cue is weighted equally and combines in a multiplicative fashion. Our data support this principle in the context of sentence parsing because we observed effects of semantic cue matches with the verbal retrieval probe even when the syntactic roles of that item suggest that it would not make an appropriate subject filler. Moreover, we found that this effect cannot be the result of a "snatch adjacent items" strategy, because the effect was observed even when additional items were added between the semantic match and the verbal retrieval probe. Thus, the best account for these results is of retrieval of recent items that match the retrieval cues on at least one dimension, despite mismatches on others. A match on all dimensions does not necessitate a correct retrieval because the matching process also is affected by recency and the retrieval strengths of all other cues associated with the target item, any of which may weaken the probability of retrieving the correct item. This was consistently illustrated with interactions between high levels of syntactic and semantic interference in our data, giving strong support to the prediction of multiplicative cue combining and attesting to the relevance of interference theory for sentence processing.

In conclusion, the data presented here clearly support a model of parsing that operates according to the principles of cue-based retrieval and global matching. In addition, we saw some preliminary evidence that establishing a grammatical dependency via cue-based retrieval is only the initial step in attaining complete comprehension, and that the creation of a situation model follows closely on the heels of establishing the dependency. The latter processes appear to be unaffected by the properties which affect global-matching, primarily because retrievals associated with elaborating the situation model are not cue-based, as there is no retrieval cue to initiate them. Rather these processes appear to be sensitive to a more general decay of the items being interpreted. Experiments 3-5 will investigate these processes further, with the goal of elucidating interactions between the cue-based retrieval processing and those associated with constructing the situation model.

3 Referential interference

3.1 Introduction

Chapter 2 presented evidence in support of retrieval effects in sentence parsing that could not be explained by storage-based models of parsing breakdown because storage requirements in the structures investigated were kept constant. In this Chapter, I manipulate storage requirements directly by using anaphoric reference to already established items in the discourse and show that in cases were storage requirements are reduced, processing is still difficult due to the necessity of retrievals. Moreover, I will illustrate that retrievals themselves can increase the amount of interference in a sentence, making correct items more difficult to access because previously retrieved items have become more active as a result of the retrieval.

3.2 Anaphoric processing

3.2.1 Storage effects

Anaphoric reference provides a method for reducing the storage requirements associated with the sentences examined in Chapter 2 while maintaining the distinction between low and high syntactic interference. For example, the low and high interference sentences in Example 3.1 and Example 3.2 require maintenance of three NPs, *worker*, *boy*, and *warehouse*.

Example 3.1: The worker was surprised that the boy who was living near the warehouse was complaining about the investigation.

Example 3.2: The worker was surprised that the boy who said that the warehouse was dangerous was complaining about the investigation.

The storage demand for processing the intervening region can be reduced if the innermost NP is replaced with pronominal, as in Example 3.3 and Example 3.4, making the creation and storage of an additional referent unnecessary.

Example 3.3: The worker was surprised that the boy who was living near her was complaining about the investigation.

Example 3.4: The worker was surprised that the boy who said that she was dangerous was complaining about the investigation.

A number of empirical studies support the view that referring to a previously established referent is less costly than creating an entirely new referent in the situation model (Garrod & Sanford, 1977; Haviland & Clark, 1974; Murphy, 1984). For example, Murphy (1984) found that reading times for the second sentence of the pair presented in Example 3.5 were slower when the NP was indefinite than for when it was introduced by the definite article.

Example 3.5: Though driving 55, Steve was passed by a truck.

Later, George was passed by a/the truck, too.

They concluded that the indefinite article signaled that the NP *a truck* in the second sentence refers not to the already mentioned truck, but to a new truck, thereby requiring the creation of a new discourse referent. In contrast, the definite article simply signaled reference to the already mentioned truck, which is a less costly operation. Similarly, they found that replacing *a truck... a/the truck* sequence in Example 3.5 with a pronominal reference sequence, as in Example 3.6 produced the same result, suggesting that creating referents is costly while referring back to them is not.

Example 3.6: Though driving 55, Steve was passed by a truck.

Later, George was passed by it/one too.

Notably, this conclusion appears to hold even when the antecedent of a referring expression is not explicitly stated in the previous context. For example, Garrod and Sanford (1982) found that reading Example 3.9 after the sentence in Example 3.8 takes no longer than reading it after Example 3.7.

Example 3.7: Keith took his car to London.

Example 3.8: Keith drove to London.

Example 3.9: The car kept overheating.

They suggested that the absence of an explicitly mentioned antecedent for *the car* in Example 3.9 was not detrimental when read after Example 3.8 because Example 3.8 facilitated the inference that Keith drove a car to London, and hence the antecedent is likely to exist in the situation model even though it was not mentioned in the text. Thus, a storage account predicts that Example 3.3 and Example 3.4 will be easier to process than Example 3.1 and Example 3.2.

3.2.2 Retrieval effects

In contrast to the storage account discussed in Section 3.2.1, a retrieval account makes the opposite predictions. This is because interpretation of the pronoun *she* requires reactivation of its antecedent in working memory. There is a considerable body of empirical evidence in support of this (e.g., (Almor, 1999; Bower & Rinck, 2001; Clark & Sengul, 1979; Dell, McKoon, & Ratcliff, 1983; Garrod & Sanford, 1977; Gernsbacher, 1989; McKoon & Ratcliff, 1980; Nordlie, Dopkins, & Johnson, 2001; O'Brien, Albrecht, Hakala, & Rizella, 1995; O'Brien, Duffy, & Myers, 1986; Sanford, Moar, & Garrod, 1988)). One particularly well-controlled study by O'Brien et al. (1986) tested for antecedent reactivation using the materials in Table 3.1.

Table 3.1: Materials from O'Brien et al. (1986)

Condition	Passage with alternative final lines
	Bill was late for work again. He had forgotten to set his alarm and had overslept. To top it all off, he was now behind a *bus* that was having engine trouble. As the bus pulled into a busy intersection, its engine sputtered and died. Bill shook his head in disbelief as the traffic around him came to a halt. There was an important sales meeting this morning, and he was afraid he would miss it. His boss had already warned him once about his tendency to oversleep. Bill sighed as he caught sight of a digital clock in a nearby bank window. He sat and stared at
Reinstatement	the stalled vehicle not knowing what to do.
Semantic control	a police vehicle not knowing what to do.
Baseline control	the digital clock not knowing what to do.
Preanaphor control	Last line deleted.

Subjects read the passage one line at a time and then were presented with a row of Xs for 300ms followed by the probe word (i.e. *bus* in the example in Table 3.1) and asked to name this word aloud. O'Brien et al. predicted that if the referring expression *the stalled vehicle* reinstated its antecedent, then naming times for the probe should be decreased relative to either of the three control conditions. In the semantic control condition, the NP *a police vehicle* allows a test of whether any facilitation of *bus* after the occurrence of *the stalled vehicle* is due simply to spreading activation among related concepts and not to the actual reactivation of the antecedent. In the baseline control condition, the NP *the digital clock* also requires reinstatement of a previously occurring concept, but the naming times for *bus* in the reinstatement condition should be faster than when this probe follows the baseline completion sentence since it is expected to be reactivated only by its own anaphor. The final control condition tests whether the antecedent has in fact become less accessible in memory because there has been no anaphoric reference to reinstate it since its original mention. If this is correct, naming times in this condition should be much greater as compared to the reinstatement condition, and a strong argument for the decay and reinstatement of antecedents can be made. Indeed, the pattern of results observed confirmed these predictions. Naming times for the reinstatement condition were significantly faster than any of the other three conditions, suggesting that the referring expression reactivates its antecedent.

This reactivation has the same properties as a cue-based retrieval, since its success is determined by how well the anaphoric expression selects the proper antecedent among the known referents in the situation model (Ariel, 1990; Ariel, 1994). This depends on two factors: the semantic retrieval cues associated with the anaphor itself (i.e., interitem associative cues with strength *b* according the language of the SAM model), and the degree of accessibility, or self-strength, of the antecedent being retrieved (i.e., cues of strength *c* in the SAM model). If an antecedent is more accessible then a pronominal anaphor can be used because this type of anaphor provides little content information to guide the retrieval. In contrast, if an antecedent is less accessible, more information is required to reinstate it and therefore a fuller referring expression (e.g., a definite NP) must be used.

Ariel provides evidence in support of this account from a corpus study of the distribution of the various types of referring expressions and their location with respect to their antecedent. Table 3.2 summarizes a subset of her data presented in Ariel (1990), showing the percentage of occurrences for each form when its antecedent occurs in each of four progressively distant locations.

Table 3.2: The distribution of referring expressions in context, data from Ariel (1990).

	Location of Antecedent (Percent of total occurrences)				
Referring Expression	Same Sentence	Previous sentence	Further back in same paragraph	Across paragraphs	Total
Pronouns	20.8	60.5	14.2	4.5	100
Demonstrative pronoun	4.8	59.5	20.2	15.5	100
First Name	6.1	42.9	40.8	10.2	100
Definite NP	2.8	14.1	45.8	37.3	100

Thus, pronouns were found when the antecedent occurs within the two minimum distances (over 81%), demonstratives and names favor the two intermediate distances (nearly 80% and over 83%, respectively), and definite descriptions favor the maximal distances (over 83%). These data seem to suggest that the use of a particular anaphoric form depends not just on whether referents can be uniquely identified, but whether they are still active in the discourse context. When sufficient distance between an antecedent and a referring expression exists, the form of that expression changes in such a way as to make it more readily identified.

Additional evidence for regarding anaphors as retrieval cues is available from an empirical study by Almor (1999), who manipulated the amount of semantic information contained in the anaphoric reference, and hence the usefulness of the anaphor as a retrieval cue. She reasoned that the more specific the anaphoric expression, the more information it contained that could be marshaled for cueing the retrieval. Thus, she compared the four conditions in Example 3.10-Example 3.13 in which the anaphoric NP in the second sentence was increasingly more specific.

Example 3.10: A robin ate the fruit. It seemed very satisfied.

Example 3.11: A robin ate the fruit. The bird seemed very satisfied.

Example 3.12: A robin ate the fruit. The wet little bird seemed very satisfied.

Example 3.13: A robin ate the fruit. The robin seemed very satisfied.

In order to evaluate the effect of the antecedent's accessibility on the usefulness of the content cues, she also manipulated whether or not the antecedent was in focus, and hence had a high self-strength making it more accessible. This was done by changing the form of the initial sentence so that either the antecedent was in focus (i.e. "It was a robin that ate the fruit.") or the object was in focus (i.e. "What the robin ate was the apple.").

She found that when the anaphor was less specific, it was read more quickly when its antecedent was in focus, suggesting that the increased availability of the antecedent compensated for a semantically impoverished anaphor. However, when the antecedent was focused and the anaphor was highly specific, reading times were slowed. Almor suggests this is due to the use of more specific anaphors than functionally necessary, thus exacting a processing cost for interpreting the new and unnecessary information in addition to identifying the antecedent. In contrast, when an antecedent was not focused, these more specific anaphors were read more quickly, supporting the view that the additional semantic information they provide was helpful, or even necessary, for identifying the appropriate antecedent.

3.3 Experiment 3: Initial evidence for retrieval effects

Section 3.2 illustrated the contradicting predictions made by the retrieval and storage accounts regarding whether the pronominal status of the inner NP should help or hurt processing. No previous empirical work has tested these predictions using 3rd person pronouns in structures such as those in Example 3.3 and Example 3.4; however, two experiments reported by Warren and Gibson (2000; in press) provide relevant data. In an off-line complexity judgment task, they compared center-embedded sentences such as those in Table 3.3, which differ in the referential properties of the most embedded NP.

Table 3.3: Referential Conditions in Warren and Gibson (in press)

Condition	Sentence
Indexical pronoun	The student who the professor who I collaborated with had advised copied the article.
Short Name	The student who the professor who Jen collaborated with had advised copied the article.
Full NP	The student who the professor who the scientist collaborated with had advised copied the article.
No Referent	The student who the professor who they collaborated with had advised copied the article.

Judging the ease of understanding each of these sentences on a 5-point scale, subjects rated the indexical pronoun (e.g., *I* or *you*) sentences as being significantly easier than any of the other sentence types, among which there were no differences in rating scores. Warren and Gibson interpret this as evidence that the parser builds new discourse entities incrementally during processing, and that each new entity the parser must create increases processing costs because storage of that entity depletes the resources available for parsing (Gibson, 1998; Gibson, 2000). In the case of indexical pronouns, they claim that no new entity must be created because referents for the speaker/writer and hearer/reader are always assumed to be part of deictic frame of the current discourse (Chafe, 1987; Enç, 1983). Consequently, the indexical conditions are easier to process than any of the other 3.

In an on-line reading time study designed to follow up this result, Warren and Gibson (in press) compared relative clause sentences as in Example 3.14 and Example 3.15 and complement clause structures as in Example 3.16 and Example 3.17, each with either an indexical pronoun *you* or a full NP *the boy*. Relative clause sentences are predicted to be more sensitive to the referential status of the embedded NP because it occurs in between the subject and verb of the relative clause.

Example 3.14: The woman who you had accidentally pushed off the sidewalk got upset and decided to report the incident to the policeman standing nearby.

Example 3.15: The woman who the boy had accidentally pushed off the sidewalk got upset and decided to report the incident to the policeman standing nearby.

Example 3.16: The woman knew that you had accidentally pushed the girl but gave you a long lecture anyway.

Example 3.17: The woman knew that the boy had accidentally pushed the girl but gave him a long lecture anyway.

Their results replicated those found with the off-line data, in that indexical pronouns were significantly easier than full NPs for the relative clause sentences, but not for the complement clause sentences. Warren and Gibson attempt to extend this finding further by importing Gundel's (1993) Givenness hierarchy into their predictions about the types of NPs that should save processing resources. Specifically, based on Gundel, they predict the order of difficulty shown in Example 3.18, as determined by savings in work necessary to activate a referent in the existing discourse.

Example 3.18: Indexical NP < Famous NP < full definite NP < full indefinite NP

I/you < NBC < the reporter < a reporter

Measuring reading times for sentences such as those in Example 3.19, they found results that are important for interpreting the claims of Gibson's (1998) storage model.

Example 3.19: The writer who {I, NBC, the/a reporter} talked to wrote radical articles about the government.

Reading times at the verbal *talked to* followed the pattern predicted by the hierarchy, except that there were no differences between the pronominal and the famous NP conditions. Consequently, they conclude that accessibility, as defined by Gundel's Givenness hierarchy, affects long distance attachment such that items that are *given,* and therefore already established in the discourse context, are less costly than those that are *new.*

I suggest that this conclusion should be questioned on two grounds. First, the materials used to test the hypothesis do not afford a clear test of their storage hypothesis. Gibson intends his model to provide a general explanation for complexity effects, including long distance attachments; however, all the evidence presented in favor of his model has included object-extracted relative clauses in the region of highest difficulty (Gibson, 1998; Kaan et al., 2000; Warren & Gibson, 2000). The result is a confound of complexity due to the long-distance attachment itself and complexity due to maintaining a long distance filler-gap dependency (Hawkins, 1999). Indeed, Gibson refers to the process of attaching items into the existing phrase marker as *integration*, a process that in his examples includes not only attachment, but also creation of object traces and binding them to their antecedent. I suggest that creation of referents may appear more important in his data because two retrievals are necessary to complete the filler-gap dependency, and difficulty is reduced when one of the items is a pronoun because this serves to distinguish the two antecedents being retrieved (see (Gordon, Hendrick, & Johnson, 2001) for a similar approach). Two indications of the correctness of this interpretation is available in Warren and Gibson's own data. First, no effects of the indexical pronoun were found in the complement structures in Example 3.16 and Example 3.17, which do not include a filler-gap dependency at the point of integration. Secondly, no differences between the referential conditions were observed at the word *wrote* in Example 3.19, which is the point where the long distance attachment between *the writer* and its verb must be made. Rather, differences were observed at the previous verbal *talked to*, which is where the filler-gap dependency is resolved.

The second criticism of the Warren and Gibson approach stems from the interpretation of *given* items as active, and hence more accessible. The problem with the association of "givenness" with "availability" is that "givenness" is a property of a discourse, while "availability" is a property of memory capacity. We saw in Section 3.2 that the storage account of referential costs makes predictions opposite those of a retrieval account. Specifically, items which do not need to be created are cheap when cost depends on memory capacity, but these same items may be relatively inaccessible to the retrieval mechanism despite being previously established in the situation model. This is because the ability to linguistically refer to a *given* item using deictic devices such as *this, the, he, she, you,* and *I* does not mean that item is active in memory. For example, although the indexical pronouns in the Warren and Gibson data are known and therefore may not need to be created, they may not be accessible from a working memory point of view because they have never been explicitly used in the current discourse. Thus, contrary to the Warren and Gibson prediction, they will be costly to process because their availability is low. Similarly, items that were mentioned long ago in the discourse, may have lower availability than those mentioned more recently, despite both having the status of *given.* Experiments 3 and 4 are designed to provide data to distinguish the storage vs. retrieval accounts in sentences requiring long distance attachments that do not also contain filler-gap dependencies.

3.3.1 Method

Following the method of Warren and Gibson (in press) I collected complexity ratings in an offline study as an initial evaluation of the hypothesis that referential status reflects antecedent availability rather than creation costs. I used high and low syntactic interference structures such as those in Chapter 2, and varied the referential status of the embedded NP so that the NP was either implicitly *given* (indexical pronoun), explicitly *given* (pronominal), or *new*. According to the storage account, both of the *given* conditions are predicted to be low cost because the antecedent for the pronoun already exists in the discourse. In contrast, the full NP condition is predicted to be difficult on this view because the NP that has not occurred previously must be created as well as integrated into the current situation model. In addition, the Discourse Locality Theory (Gibson, 1998; Gibson, 2000) suggests a possible interaction between referential status and sentence type, since the two Given conditions are expected to make the high interference sentence less difficult because the intervening NP need not be created. This savings of resources may make these conditions comparable to the low interference conditions which contain an NP which must be created, but since it is not embedded in an additional clause it will not be as difficult as the NP condition for the high interference sentences. Finally, the least complex sentences are expected to be the low interference pronominal and indexical conditions, since they contain *given* NPs and no additional embedding.

3.3.1.1 Subjects

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

3.3.1.2 Materials

Thirty-six experimental items were created, each having six conditions, as presented in Table 3.4. Three versions of high and low syntactic interference sentences such as those used in Experiments 1-2 were created by manipulating the referential status of the inner-most NP- the one predicted to be the most costly because it occurs in the region between the long distance dependency. An indexical condition was created by inserting either *I* or *you* as this NP, with half the items containing each pronoun. A pronominal condition was created in which the embedded pronoun unambiguously refers back to the subject of the sentence. In cases where it was not possible to distinguish the prior NPs by gender, grammatical number was used.

Table 3.4: Materials for complexity judgments, Experiment 3

Condition	Low interference	High interference
Indexical (Implicit Given)	The young girl saw that the boy who was sitting next to me was asleep before the crash.	The young girl saw that the boy who noticed I was uncomfortable was asleep before the crash.
Pronominal (Explicit Given)	The young girl saw that the boy who was sitting next to her was asleep before the crash.	The young girl saw that the boy who noticed she was uncomfortable was asleep before the crash.
Full NP (New)	The young girl saw that the boy who was sitting near the door was asleep before the crash.	The young girl saw that the boy who noticed the seat was uncomfortable was asleep before the crash.

Six experimental forms were created so that each subject saw only one condition for each set of items. The experimental sentences were randomly interspersed among 39 filler items, which were designed to be suitable comparison sentences. In order to match the structures of our interfering items, one third of these 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 the low interference items, except that no long distance attachment was required. Another third of the filler items 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 third 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 the complexity rating.

Subjects were presented with a packet containing the 75 sentences and asked to rate each one according to the scale in Example 3.20, which was reprinted in the packet following each sentence. The scale annotations were included in order to control for a possible flaw in the original Warren and Gibson (in press) complexity study: Without such annotations, subject responses may not be comparable since individual subjects may interpret the scale differently.

Example 3.20

1-Not complex, easy to understand.

2-A little complex, but generally understandable.

3-Rather complex, but I could understand it after some thought.

4-Complex, and I understood it only somewhat.

5-Very complex, hard to understand.

3.3.1.3 Procedure

The experiment was conducted in a large lecture hall, with all subjects simultaneously completing a printed packet containing the materials described above. The cover page of the packet instructed subjects to make their ratings on a scantron bubble sheet that was later scored by the University of Pittsburgh

Office of Measurement. They were given as much time as necessary to complete the packet, and instructed to make their best guess on any item they were unsure of.

3.3.2 Results & Discussion

The subjects' complexity ratings were submitted to a 2 (sentence type) x 3 (referent type) ANOVA. The main effect of sentence type was significant, $F_1(1,48) = 75.57$, $p < .001$; $F_2(1,35) = 34.22$, $p < .001$, with the high syntactic interfering sentences being more difficult than the low syntactic interference structures. Likewise, the main effect of referent type was significant, $F_1(1,48) = 22.91$, $p < .001$; $F_2(1,35) = 9.12$, $p < .001$. This effect was due to a large difference between the indexical condition on the one hand, and the pronominal and NP conditions on the other hand. The interaction was not significant. Table 3.5 presents the mean ratings in each condition.

Table 3.5: Means and (Standard Error) for Complexity Ratings in Experiment 3.

Referent type	Low syntactic interference	High syntactic interference
Indexical	2.08 (.10)	2.43 (.10)
Pronominal	1.84 (.09)	2.13 (.10)
NP	1.76 (.09)	2.21 (.10)

Contrary to the prediction of Warren and Gibson, these results indicate that an indexical pronoun does not decrease sentence complexity in our sentences, but rather increases it. If the assumption that the indexical already exists in the discourse is correct, then this must be due to the fact that the referent must nevertheless be activated, since it has not been explicitly mentioned previously. If this assumption is not correct, then the costs associated with indexical NPs appear to be even greater than those associated with a full NP, perhaps deictic properties of the pronoun initiate an attempt to reactivate the presupposed referent, which has, in fact, not occurred. In either case, the important cost associated with these sentences comes from the ease with which the antecedent of the indexical pronoun can be accessed. In contrast, the antecedent for the pronominal condition was explicitly mentioned in the sentence, and can be unambiguously identified by the anaphor. This makes the reinstatement easy compared to the indexical condition, contrary to the Warren and Gibson account, which would predict that both pronominal (i.e., *given*) conditions should decrease processing equally. Together, these results support an account that measures processing costs not along the lines of whether antecedents are known in the current discourse, but whether they can be easily accessed. In the full NP condition, access is restricted only by the interference from other items in the intervening region. The pronominal conditions did not differ from the full NP conditions, as should be the case because these conditions were identical in the amount of structural interference they contained. The only additional processing cost to be expected would have come from difficulties interpreting the pronominal, and since the sentences were constructed to make the pronominal antecedent easy to identify, this cost should have been minimal.

3.4 Experiment 4: More evidence for retrieval effects

While Experiment 3 provided significant effects, the use of the offline complexity measure contains a high degree of uncertainty regarding the criteria subjects are using to make their judgments. Consequently I attempted to replicate the results of Experiment 3 with a grammaticality judgment task, which provides a well-defined measure of complexity effects. Accuracy rates in each condition will provide an evaluation of how often subjects were unable to interpret the sentences in that condition, which reflects the complexity of each sentence type.

In addition, it was discovered that roughly 25% of the high interference materials in Experiment 3 contained an ambiguity in the embedded clause. For example, in Example 3.21 the embedded verb *noticed* could take either an NP object or a clausal object.

Example 3.21: The young girl saw that the boy who noticed the seat was uncomfortable was asleep before the crash.

In leaving out a *that* in the embedded clause, this created potential for additional complexity, and while this was not a problem for low interference sentences, it may have influenced processing on the high interference sentences in Experiment 3, inflating the effect of sentence type in that experiment. To correct for this, Experiment 4 used the same materials as Experiment 3, but with a *that* inserted into the embedded clause in all the high interference sentences.

3.4.1 Method

3.4.1.1 Subjects

Thirty-six University of Pittsburgh undergraduates participated in the experiment for partial course credit. All subjects were native speakers of American English with vision corrected to normal. Due to a technical difficulty with the computers used to present the experiment, 7 subjects' data was lost, so analyses were performed on data from only 29 subjects.

3.4.1.2 Materials

The experimental materials in this experiment were identical to those in Experiment 3 except for the changes already noted. In addition, since subjects were making grammaticality judgments, we included an additional 36 filler items that were ungrammatical due to the addition or deletion of NPs or verbs.

3.4.1.3 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 be either missing words or have extra 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 commas and that subjects should not use missing punctuation as a reason for judging a sentence to be 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.

3.4.2 Results and Discussion

Accuracy scores were submitted to a 2 (sentence type) x 3 (referent type) ANOVA. Table 3.6 shows means and standard error scores for each experimental condition. The main effect of sentence type was significant, $F_1(1,28) = 4.42$, $p < .05$; $F_2(1,36) = 6.93$, $p < .02$, with the high interference sentences being more difficult than the low interference sentences. The main effect of referent type was not significant, $F_1(1,28) = 1.96$, $p =.15$; $F_2(1,36) = 2.80$, $p =.07$, however, the interaction of sentence and referent types was significant, $F_1(1,28) = 3.78$, $p < .03$; $F_2(1,36) = 3.15$, $p < .05$. To understand this interaction more clearly, separate analyses were done on data for each of the two sentence types. For the low interference sentences, the effect of referent type was significant, $F_1(2,56) = 3.88$, $p < .03$; $F_2(2,70) = 4.62$, $p < .02$. This effect was due primarily to a significant difference between the easier NP condition and either of the other two. However, The effect of referent type disappeared for the high interference sentences, $F_1(2,56) = 1.93$, $p = .16$; $F_2(2,70) = 1.93$, $p = .15$, as none of the pairwise comparisons were significant.

Table 3.6: Accuracy scores (standard error) for Experiment 4

Condition	Low interference	High interference
Indexical	.81 (.05)	.74 (.05)
Pronominal	.81 (.04)	.82 (.03)
Full NP	.91 (.03)	.75 (.04)

The Discourse Locality Theory cannot explain this interaction, as the Given-New status of referents is expected to have an effect in both sentence types. Moreover, the indexical conditions were predicted to be easier than the full NP conditions on that account, yet the opposite pattern was once again observed here.

These data can best be accounted for by a retrieval account of processing costs. To make this clear, I first consider the low syntactic interference sentences, in which the referent type effect was found. The presence of this effect indicates that the referential status manipulation affected the amount of interference created by the inner NP in these conditions, with greater difficulty resulting from greater interference. To understand this, we must examine the probability of retrieving each NP in the sentence, again making use of the SAM model's equation for estimating retrieval probabilities, repeated here as Equation 3.1.

$$P(I_i|Q_1, \ldots, Q_m) = \frac{\prod_{j=1}^{m} S(Q_j, I_j)^{w_j}}{\sum_{k=1}^{N} \prod_{j=1}^{m} S(Q_j, I_k)^{w_j}}$$

Equation 3.1: SAM retrieval function

Table 3.7 illustrates estimated retrieval strengths for each NP in the low interference conditions using the items in Table 3.4 to provide concreteness to the example. As in previous examples, values for syntactic cues have been estimated by assessing how well the features of the NP match the retrieval cues of the verb; the exact values are less important than the relative differences between the conditions. For example, since the verb is looking for a subject NP, both the pronominal and indexical anaphors have low retrieval strengths with respect to these cues because they are both case-marked as syntactic objects. Similarly, although the full NP is not case marked, it nevertheless carries a grammatical case, which gives it syntactic features that do not match with those of the verb. Its retrieval strength may be higher than that of the pronominals because the lack of overt case may make it a slightly better match, but it is still quite a low match compared to items which are explicitly case-marked as subjects.

Table 3.7: Retrieval strengths for intervening NPs in low syntactic interference conditions

Condition	Linguistic item	Syntactic feature match	Self-strength	Estimated probability of retrieval (numerator only)[6]
Full NP	door	.12	.98	.12
Pronoun	her	.02	.98	.02
Indexical	me	.02	.98	.01
Pronominal	girl (antecedent)	.95	.79	.77
Indexical	I (antecedent)	.98	.54	.53

The pronominal and indexical conditions appear twice in this table, because the fact that they initiate reinstatement of their antecedents means that these items must also be considered as potential sources of interference. Self-strength values were assigned according to the recency of items vis-à-vis the verbal retrieval probe: since items in the intervening region occurred immediately prior to the verb, their self

[6] Note from Equation 3.1 that the probability of retrieving a particular item is a ratio of the cue-to-image strength for that item and the sum of the cue-to-image strengths for all items. Since the examples discussed here are identical except for the NPs depicted in Table 3.7, it is only the cue-to-image strength for this NP that matters for a comparison of these conditions, and hence only these values are presented.

strength was extremely high. In the case of antecedents, I averaged an estimate of its self-strength prior to reinstatement together with the self-strength of its anaphor (cf. Hintzman's (1988) MINERVA 2 memory model, which suggests that self-strength is the mean of all occurrences of that item in memory.)[7]

From Table 3.7, it is clear that the two antecedents supply a large amount of interference, resulting from their reactivation and their close syntactic match with the features sought by the verb. Specifically, the antecedent of the pronoun is a subject, so it might appear to be exactly what the verb is looking for, hence giving it a high retrieval probability. The same is true for the indexical, primarily because having not been mentioned before, it has no grammatical role associated with it and may look even more appropriate for that reason. This is supported by the fact that the effect of syntactic interference disappears for both the pronominal, $t_1(28) = -.191$, $p = .85$; $t_2(35) = -.27$, $p = .79$, and indexical conditions, $t_1(28) = 1.31$, $p = .20$; $t_2(35) = 1.51$, $p = .14$, suggesting that the syntactic marking on the anaphor itself is irrelevant to reducing syntactic interference, since its source is from the antecedent, whose case marking makes it interfering even in the low interference conditions. [8] In contrast, the NP conditions do not reactivate antecedents, so the syntactic interference effect generated from the different grammatical case for this NP in the low and high interfering conditions remains, $t_1(28) = 3.21$, $p < .004$; $t_2(25) = 3.52$, $p < .002$.

Turning now to the high interference sentences in which the effect of referent type disappears, I suggest an account similar to that described above, in which retrieval results in reactivation of referents, and this reactivation increases the amount of interference contributed by those items. To illustrate, we consider the retrieval points identified with the subscripted numbers in the sentences in Table 3.8.

---

[7] Since the pronominal antecedent occurred several words prior to its anaphor, I assumed that this antecedent had decayed to the level of .60. To obtain its self-strength after reactivation, I averaged .60 and .98. For the indexical antecedent, since it never occurred in the discourse, I averaged .10 and .98, with the .10 reflecting the possibility that the antecedent was understood despite not being mentioned.

[8] The suggestion that properties of the antecedent can contribute interference for the grammatical dependency in these sentences is particularly interesting because the antecedent does not actually occur in the intervening region. Consequently, this provides important evidence for a retrieval account of anaphor processing, since its interference is the direct result of antecedent reactivation. Further experiments investigating properties of the antecedent which contribute interference in structures such as these, including frequency and semantic properties, are planned.

Table 3.8: Retrieval points for high interference sentences.

Condition	Sentence
Indexical	The young girl saw that the boy who noticed that I_1 was_2 uncomfortable was_3 asleep before the crash.
Pronominal	The young girl saw that the boy who noticed that she_1 was_2 uncomfortable was_3 asleep before the crash
Full NP	The young girl saw that the boy who noticed that the $seat_1$ was_2 uncomfortable was_3 asleep before the crash.

The SAM model discussed in Section 2.2 predicts that repeated retrievals increases an item's self-strength because they provide additional rehearsal opportunity for the items retrieved (see also Hintzman's (1988)). Consequently, by the time it is necessary to retrieve a subject at retrieval point 3, the intermediate subjects in all three sentences have become highly active by the retrievals at points 1 and 2. In the case of the full NP, point 1 is not actually a retrieval point, however, the item has its highest self-strength at this point because it was just created. In the case of the pronominal condition, the strength of the pronominal antecedent is boosted by the use of the anaphor (recall that the new self-strength will be averaged with that from its previous mention) so that although it may still not have the activation strength of the full NP at this point, it nevertheless provides more interference than if it had not been reinstated. Similarly, in the case of the indexical, this item has an increased self-strength because of its explicit mention in the discourse. In order to make this more concrete, Table 3.9 illustrates these comparisons with numerical estimates of each item's self-strength.

Table 3.9: Retrievals for high interference sentences in Table 21.

	Linguistic item	Retrieval point 1	Retrieval point 2
Full NP	door	.98	.98
Pronominal/antecedent	she/girl	.82	.90
Indexical/antecedent	I/I	.75	.87

Thus, because the syntactic structure of the high interfering sentences includes an additional retrieval in the intervening region, the pronominal conditions provide nearly as much interference as does the full NP condition, causing the effect of referent type to disappear. This means that the critical factor is not whether or not the antecedent has been previously mentioned, but rather how active it is in working memory. Experiment 5 presents additional online data in support of this account.

3.5 Experiment 5: Online measures

Although I have argued that the effects observed in Experiments 3 and 4 are retrieval effects, I have not reported evidence directly supporting this conclusion because the dependent measures in those experiments were taken well after the retrieval is thought to have occurred. Consequently, Experiment 5 reports reading times for the regions in which the retrievals are expected to occur. In addition, I have argued that the referent type effect observed for low interference sentences in Experiment 4 is a function of the anaphoric reinstatement of a previous referent. Nevertheless, it may still be the case that the difficulty associated with these sentences was not due to interference from the reinstated referent, but rather from processing of the pronoun- either in the pronominal or in the indexical conditions. Data from the study by Almor (1999) discussed in Section 3.2.2 suggested that pronouns were most easily processed when their antecedents were in focus. This was not the case in either the low interference or the high interference conditions, so that despite the fact that the pronoun unambiguously identified its antecedent, it may still have been costly to process. In fact, it may have been this cost, which would have been necessary for both sentences types, which accounted for the diminished effect of sentence type in the pronominal conditions. For this reason, Experiment 5 manipulates the given/new status of a referent by reinstating that referent with the same lexical form, making any additional processing related to interpreting the pronoun unnecessary. For the sake of comparison with a pronominal condition, and in order to test once again the prediction that indexical NPs reduce processing load, we include the indexical NP condition here as well.

3.5.1 Method

3.5.1.1 Subjects

Thirty-six undergraduates from the University of Pittsburgh participated in the experiment for partial course credit. All were native speakers of American English with vision correct to normal.

3.5.1.2 Materials

Fifty items from the pool described in Section 2.4 were chosen for use in this experiment. As in Experiments 3 and 4, only the high and low syntactically interfering versions were used, since we are concerned in this experiment only with how referential status affects cue-based retrieval. Each sentence was manipulated so that the referential status of the inner most (i.e. interfering) NP will have a different initial activation based on its *given* or *new* status (See Table 3.10). For the *new* conditions, the identical form of the sentence as described in Section 2.2 was used. To create the "old" condition, the identical sentence was used, except a clause containing the inner NP was appended to the beginning of the sentence so that at the point of cue-based retrieval this NP is considered *given*. A final condition using the indexical pronoun *you* was included in order to replicate the effect seen in Experiments 3 and 4. The findings there suggest that this NP will act like a New NP since it has not been established in the discourse prior to its occurrence in the region between the target NP and the verbal retrieval probe.

Table 3.10: Example Referential interference stimuli

	Low syntactic Interference	High syntactic Interference
New	The // assistant forgot that the client / who had asked about the meeting / was waiting / for a // signature.	The // assistant forgot that the client / who implied that the meeting was important / was waiting / for a // signature.
Given	The meeting was ready to begin, and the // assistant forgot that the client / who had asked about the meeting / was waiting / for a // signature.	The meeting was ready to begin, and the // assistant forgot that the client / who implied that the meeting was important / was waiting / for a // signature.
Indexical	The // assistant forgot that the client / who had asked about you / was waiting / for a // signature.	The // assistant forgot that the client / who implied that you are important / was waiting / for a // signature.

3.5.1.3 Procedure

The procedure for this experiment is the same as that employed in Experiment 2, which was described in Section 2.6.1.3.

3.5.1.4 Dependent measures and Analysis

Due to an error in materials preparation, the low interference sentences were queried with comprehension questions about the target NP, while the high interference sentences were queried with questions regarding the matrix subject. This split makes analyses of the comprehension questions uninteresting, and so those data are not reported here.

Reading time data was analyzed using the same procedure as described in Section 2.6.1.3. As with that experiment, data points that were either too fast (150 ms) or too slow (2 seconds) were eliminated from the analysis. In addition residual reading times that were either 3 times the interquartile range above the upper quartile or below the lower quartile were replaced with the cutoff value. These procedures affected approximately 2% of the data points.

Following the procedure described in Section 2.6.1.3., we calculated separate regression analyses for each subject, with word position and word length as predictors. The regression on word position provided a correction for the fact that the critical region occurred much earlier in the new conditions than in the old conditions due to the added initial clause. Twenty-five of the 36 subjects showed negative coefficients for position in their regression equations, suggesting that there was a tendency to speed up as subjects read through the sentence. Table 3.11 presents the mean coefficients for each predictor, as well as t-tests indicating that they all account for a significant portion of the variance.

Table 3.11: Coefficients for Experiment 4 regression predictors

Predictor	Mean	SD	t-test	p-value
Constant	333.43	90.31	$t(35) = 22.15$	$p < .001$
Position	-1.61	4.34	$t(35) = -2.22$	$p < .03$
Word length	10.63	6.32	$t(35) = 10.10$	$p < .001$

Regions of interest were then created by summing over the residuals for each word in the relevant region. Regions for this experiment are the same as those used in Experiment 2, and are delimited in Table 3.10 by the single slashes. In addition to not including the first word of the sentence, as was done in Experiment 2, the initial clause in the given conditions was also not analyzed because there were no comparison conditions for this region.

3.5.2 Results & Discussion

As with Experiment 2, reading time analyses were conducted only on items for which comprehension questions were answered accurately. No subjects or items were thrown out due to missing data points, indicating that subjects were never consistently incorrect in a given condition or for a given item. Since this procedure creates different means for subject and items analyses, Table 3.12 and Table 3.13 presents the means for each analysis separately. Analyses on raw reading times were also conducted, and reported in Appendix B. The same pattern of results as reported here with the residual analyses was found, except for where variables factor out by the predictor variables influenced the results. These results are summarized in Appendix B as well.

Table 3.12: Trimmed residual reading times for each region (ms), subjects as random factor.

Low syntactic interference				Region 1	Region 2	Region 3	Region 4
	New			32.72 (28.86)	216.54 (55.21)	19.19 (15.48)	-16.29 (10.16)
	Given			-52.48 (29.37)	23.32 (41.41)	-38.19 (14.85)	38.47 (10.28)
	Indexical			-8.06 (25.74)	195.54 (66.91)	-15.47 (19.72)	-21.92 (11.11)
High syntactic interference							
	New			-7.64 (35.69)	248.32 (82.62)	59.66 (25.85)	4.39 (13.62)
	Given			-30.36 (39.05)	25.18 (54.70)	5.02 (26.46)	20.60 (16.98)
	Indexical			3.94 (31.41)	315.10 (77.21)	102.52 (33.14)	-2.82 (12.28)

Table 3.13: Trimmed residual reading times for each region (ms), items as random factor.

Low syntactic interference				Region 1	Region 2	Region 3	Region 4
	New			24.95 (46.90)	189.11 (74.85)	5.40 (22.29)	-21.10 (13.36)
	Given			-56.68 (52.30)	29.23 (55.08)	-41.98 (20.05)	35.08 (16.12)
	Indexical			-9.86 (52.34)	211.44 (64.46)	-11.24 (23.63)	-21.93 (13.86)
High syntactic interference							
	New			-8.36 (45.69)	251.23 (77.65)	57.67 (22.88)	4.34 (12.09)
	Given			-12.84 (50.50)	43.98 (65.38)	8.16 (23.38)	27.53 (14.00)
	Indexical			22.03 (54.33)	361.93 (90.46)	107.15 (29.58)	.193 (16.04)

No significant effects were found in Region 1. In Region 2, a significant effect of referential status was observed, $F_1(2,70) = 9.90$, $p < .001$; $F_2(2,98) = 18.76$, $p < .001$. This was due to the given condition being read more quickly (M_1=24.25; M_2=36.61) than either the new condition (M_1=232.43; M_2 = 220.17) or the indexical condition (M_1 = 255.32, M_2 = 286.69). This pattern of results is contrary to the prediction of Warren and Gibson (in press) regarding the status of the indexical pronoun. It appears that both the new and the indexical conditions require the creation of new referents, requiring more computational resources and hence longer reading times than the given condition, which does not require any additional referential operation.

As noted above, the retrieval interference theory does not predict differences of sentence type in region 2, whereas Gibson's Discourse Locality Theory does predict such a difference in this region. The data revealed that the effect of syntactic interference was not significant, although there was a trend towards significance, $F_1(1,35) = 3.51$, $p < .07$; $F_2\ (1,49) = 3.73$, $p < .06$. Although the interaction of referential status and syntactic interference was not significant, $Fs < 1.07$, $ps > .35$, separate comparisons of the reading times for each referential condition in both the low and high interfering sentences provides an explanation of the marginal syntactic interference effect. In particular, the effect of syntactic interference was not significant for the new or the given conditions, $ts \approx 0$, but there was a trend for significance with the indexical items, $t_1(35) = -1.73$, $p = .09$; $t_2(49) = -1.90$, $p = .06$. To understand this, recall from the discussion of Experiment 4 that the high interference sentences require an intermediate retrieval to process the embedded verb (retrieval point 2 in Table 3.8), which is not required for low interference sentences. In general, this is not predicted to be costly because there are no interfering NPs and the correct NP is directly adjacent to the verbal retrieval probe, and hence has a comparatively higher activation than any competing NPs that may exist. At the referential level, however, this retrieval results in the elaboration of the subject's discourse referent with the semantic properties associated with the embedded verb. When the referent can be easily identified, as in either the *given* or *new* conditions, no additional processing cost accrues. However, in the case of the indexical pronoun, the elaboration is more costly because the deictic nature of the pronoun directs activation to a previous discourse referent that in fact has not been established in the discourse. Hence, in order for the elaboration occur, the processor must initiate a search for a presupposed referent, and the inevitable failure can account for the increased reading times in this condition.

Region 3 is the critical region for measures of retrieval interference for completing the long distance dependency. The main effect of referential status was significant in this region, $F_1(2,70) = 4.81$, $p < .02$; $F_2\ (2,98) = 9.01$, $p < .001$. This effect is due to the *given* condition being read significantly more quickly ($M_1 = -16.59$; $M_2 = -16.91$) than either the *new* condition ($M_1 = 39.43$; $M_2 = 31.54$) or the indexical condition (M_1 = 43.52; M_2 = 47.96), suggesting that the *given* condition supplies less retrieval interference than either the *new* or the indexical conditions. This is consistent with the view that the activation of the *given* NP is lower than either of the other two conditions because it was established earlier. Thus, although it was reinstated by the second mention, its self-strength is the average of the self-strengths associated with the earlier mention as well as this one, which will cause it to be lower than if it had occurred for the first time. This low self-strength means that the probability of retrieving the item is reduced as compared to the other conditions whose self-strength will be significantly higher.

Region 3 also indicated a significant effect of syntactic interference, $F_1(1,35) = 17.00$, $p < .001$; $F_2(1,49) = 30.40$, $p < .001$. This is due to the high interference condition being slower (M_1 = 55.73; M_2 = 57.66) than the low interference conditions (M_1 = -11.49; M_2 = -15.94), which is as expected due to retrieval interference from the intervening NP. In addition, there was a trend for an interaction between referential status and syntactic interference in region 3, $F_1(2,70) = 2.72$, $p < .08$; $F_2(2,70) = 2.62$, $p < .08$. This can be understood by separate analyses of the referent type effect for each syntactic interference condition. For the low syntactic interference conditions, the effect was significant by subjects, $F_1(2,70) = 4.75$, $p < .02$, but not by items, $F_2(2,98) = 2.71$, $p < .08$. Pairwise comparisons indicate that this was due to a difference between the new and given conditions, ($p_1 < .02$; $p_2 < .06$), with other pairs being insignificant. For the high syntactic interference conditions, the effect was also significant, $F_1(2,70) = 3.68$, $p < .03$; $F_2(2,98) = 7.50$, $p < .001$, but with a different pattern. Namely, for high interference conditions, the difference between new and given conditions was not significant ($p_1 < .60$; $p_2 < .13$). Instead, the difference for interfering conditions was due to the given and indexical conditions ($p_1 < .05$; $p_2 < .003$). These results fit with the pattern observed in Experiment 4 and can also be explained by the account given there. In particular, the lower self-strength of the *given* item contributes to reducing the retrieval interference in the low interference conditions as compared with the *new* condition, which is identical to it in every other respect (i.e. no explicit case marking, etc.). However, this advantage disappears in the context of a high interference sentence because the intermediate retrieval has boosted the activation of the *given* item so that it can no longer be distinguished from the new item on the bases of self-strength alone. If we then consider the indexical conditions independently, low reading times in the low interference condition reflect the fact that the case-marked indexical supplies little interference for retrieving the appropriate subject. In contrast, it supplies much more interference in the high interference structures because the pronoun has received increased activation from prior retrievals. Moreover, it is also suffers from semantic interference which the other two conditions do not have, giving it an advantage over the intended subject and significantly increasing reading times in this region.

Region 4 is the post retrieval region where we expect to observe effects associated with constructing the situation model, and especially with integrating the newly created grammatical dependency into the existing discourse context. Since the crucial syntactic dependency was resolved in region 3, no effects of syntactic interference were expected in this region, and indeed the data bore this out. I did observe an effect of referential status, $F_1(2,70) = 10.26 = 10.26$, $p < .001$; $F_2(2,98) = 10.85$, $p < .001$, however, this effect is quite different from the effect observed in region 3. In region 4, the given condition is significantly slower ($M_1 = 29.53$; $M_2 = 31.30$) than either the new condition ($M_1 = -5.95$; $M_2 = -8.38$) or the indexical condition ($M_1 = -12.37$; $M_2 = -10.87$). I suggest that this is due to the fact that the factor that aids retrievals in region 3, actually makes the integration work occurring in region 4 more difficult. Namely, if an interfering item has low self-strength, it has a lower probability of being retrieved in the global matching process and hence retrieval interference will be reduced. Yet when items are being integrated into the situation model, if they are less active then they will be more difficult to integrate, as illustrated in the literature on suppression effects (e.g. (Gernsbacher, 1989; Wiley, Mason, & Myers, 2001)) This account is supported by a highly significant region (region 3, region 4) x referent type interaction, $F_1(2,70) = 13.85$, $p < .001$; $F_2(2,98) = 29.33$, $p < .001$ (see Table 3.14 for means). Thus, although the creation of the syntactic dependency in these sentences suffers when intermediate items are highly active, when it becomes necessary to integrate the syntactic dependency with those other items to create a coherent situation model, their high activation becomes advantageous.

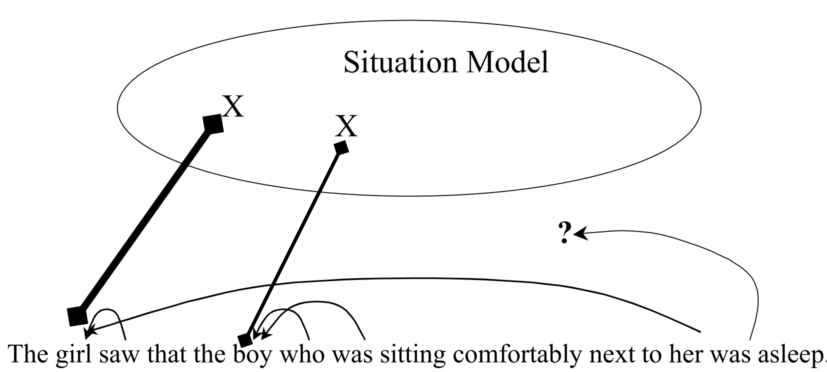
Table 3.14 Experiment 5, means for region x referent type interaction.

	Means with subjects as random factor		Means with items as random factor	
	Region 3	Region 4	Region 3	Region 4
New	39.43 (17.07)	-5.95 (10.70)	31.54 (18.42)	-8.38 (11.13)
Old	-16.59 (15.92)	29.53 (11.77)	-16.92 (19.12)	31.30 (12.82)
Indexical	43.52 (23.63)	-12.37 (10.03)	47.96 (23.93)	-10.87 (13.30)

3.6 Conclusions

Contrary to the suggestion of Warren and Gibson (in press), the experiments presented here provide no support for a storage account of referential costs, but rather suggest that referential status influences processing because it initiates retrievals of previous referents. These retrievals affect the activation of items in memory, which may cause an increase in the amount of interference for retrieving the desired grammatical filler. For example, in the sentence in Figure 3.1 the referent *girl* will have decayed considerably prior to pronominal reinstatement, yet after reinstatement it becomes even more active than the more recent NP *the boy*, which is the correct subject for *was asleep*. Consequently, completing the long distance dependency in this sentence is difficult, despite the fact that there is no syntactic interference in the intervening region.

Figure 3.1: Pronominal reference in a low interference sentence
Heavier lines indicate higher activation strength of referents.



Models that attempt to account for processing difficulty in terms of storage (e.g., Gibson (1998; 2000)) can not account for this effect because the interpolated NP has been previously established, and is therefore already stored in memory. In addition, there are no additional embeddings in the intervening region to further tax the available resources. Thus, the sole determiner of processing breakdown in this sentence is the competing self-strength of items in the discourse context, and it is this self-strength which determines how easily the correct item can be retrieved.[9]

Thus, once again the retrieval function of sentence processing is a better determiner of processing costs than metrics referring to the content of working memory. In addition, the interaction between referent type and region in Experiment 5 suggests that properties promoting the completion of grammatical dependencies via cue-based retrieval have the opposite effect on the processes responsible for integrating new information into the situation model. Chapter 4 will discuss the implications of this interaction and attempt to provide a unified model for syntactic and referential processing.

[9] There is also potential for semantic interference between *the girl* and *the boy,* because both will fit as subjects of the verbal *was asleep.* Further experimentation is necessary to determine how much of the observed effect is due to semantic interference, however since the two are most different with respect to their self-strength, it is likely that this is the most important factor in determining the probability of retrieving the correct NP in these sentences.

4 Towards an integrated retrieval model of syntactic and discourse processing

This dissertation presents data in support of the view that sentence comprehension becomes difficult when necessary items can not be efficiently retrieved from working memory. Retrieval failure can arise both from insufficient or ambiguous retrieval cues, as illustrated by the syntactic and semantic interference effects found in Chapter 2, or from insufficient self-strength for overcoming more active alternatives, as in Chapter 3. The use of the SAM retrieval model provided a concrete framework within which to examine these effects, and facilitated the generation of predictions about how and when retrieval failures might occur. For example, recall that syntactic and semantic cues are combined equally to create a composite retrieval prompt, which is then compared simultaneously to all items in working memory (cf. Equation 2.1). Because of this, I predicted that retrieval interference could occur from items which are a bad match to the retrieval prompt on syntactic grounds, but are a good match on semantic grounds. This prediction was borne out, illustrating not only the phenomenon of semantic interference, but the usefulness of the SAM model as a model of syntactic processing.

The usefulness of this model in accounting for parsing effects raises a challenge to the notion of a parser as a language-specialized structure-building device. Rather, it appears that much of the work normally ascribed to a parser (i.e., creating grammatical dependencies), can actually be described in terms of paired recognition, where the two items participating in a grammatical dependency serve as cue item and target item in much the same way that paired recognition occurs in list learning paradigms (Nobel & Shiffrin, 2001). Doing away with the distinction between parser and memory retrieval mechanism would permit making the relationship between language processes and the underlying memory systems that support it transparent, while preserving a number of properties about syntactic processing which are already known to be important. For example, a parser as retrieval approach looks very much like a head-driven parser (Kamide & Mitchell, 1999; Pollard & Sag, 1994; Pritchett, 1991; Vosse & Kempen, 2000), which builds structure only when it is licensed by a grammatical head, which on the retrieval account provides retrieval cues and initiates the retrieval. The retrieval account based on a model like SAM has the added benefit of elucidating the process by which grammatical heads are associated with their fillers. In most head-driven parsers, this is assumed to occur via a process of feature unification (Shieber, 1986), but no discussion of how, or which, features are combined is provided. Here, the process is explicitly defined because the features of the grammatical head provide the retrieval cues for identifying required fillers. In addition, the unification process of a head-driven parser is typically not susceptible to the properties of memory that make retrieval difficult (i.e. decay, interference), while we have seen considerable evidence in the experiments presented for the role of these factors in syntactic processing.

One head-driven parser that does build in some memory-related constraints is the Unification Space parser of Vosse and Kempen (2000), who assumes that the lexical frames in a fully lexicalized grammar carry a gradually decaying activation value. In this parser, each word is associated with a fully lexicalized grammatical frame, illustrated in Figure 4.1.

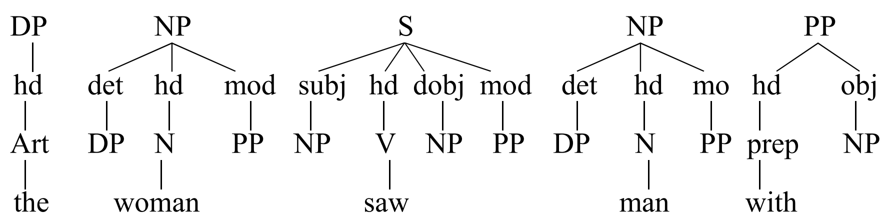


Figure 4.1: Five lexical entries for the Vosse and Kempen (2000) Unification Space parser

As each word is encountered, the lexical frame is introduced into the Unification Space, and candidate unification-links are proposed. For example in the sentence *The woman saw the man with binoculars,* when *with* is entered into the Unification Space, it will automatically be attached in the modifier position of *woman, saw,* and *man* because each of these slots can be filled by a PP, and *with* fits the bill. Unlike most unification-based parsers, however, unification in this parser occurs gradually, with each link having an associated strength, proportional to the activation levels of the two items being joined. The linking mechanism is defined so that link strength increases quickly in the absence of counteracting forces, until it reaches the value of 1 and the items are considered to be linked. However, if an item participates in more than one link, as in the current example, then these links will compete over a series of several cycles until one wins out. This competition is defined partially by constraints on tree-formation, and on word order, but also by the activation levels of the items being linked. Since item activation is assumed to gradually decrease to zero, the parser has a preference for attaching to more recent items and *with* will be attached to the NP *man*.

This parser serves as an interesting comparison for elucidating a number of ways that cue-based retrieval parsing differs from standard head-driven parsers. We will discuss these properties in the sections to follow, along with the empirical evidence relevant for distinguishing between the two accounts.

4.1 Proposing links

The architecture of a head-driven parser entails that the grammatical head initiates a search for a filler constituent to fill the roles specified by its argument structure; however, there is some debate over how this search occurs. Pritchett (1992) has suggested that the scope of the search is limited to the maximal projection of the head of the constituent, meaning that processibility is determined by abstract grammatical configuration and not simply the surface word patterns. According to this view, the parser follows a sequence of grammatical dependencies, which (hopefully) lead to the appropriate filler (see also (Fodor & Inoue, 1994; Fodor & Inoue, 2000)). In contrast, the Unification Space parser described above attaches an item to all possible attachment sites *in parallel*, and relies on link competition to settle on the correct link. This means the search comes for free, and is not influenced by locality or existing grammatical dependencies. These two search methods can be referred to as *context-addressable* (i.e., dependent on the location or organization of items) and *content-addressable* (i.e. dependent on content of items), and can be distinguished by data on the effect of the amount of material interpolated between the initiator and the target of the search. In a context-addressable system, retrieval speed will slow down if more material must be processed before arriving at the desired item, while it will remain constant for a content-addressable system.

McElree (2000) investigated this question using a speed accuracy tradeoff procedure to derive measures of speed and accuracy in processing sentences such as those in Example 4.1-Example 4.3.

Example 4.1: This was the book that the editor admired.

Example 4.2: This was the book that the editor who the receptionist married admired.

Example 4.3: This was the book that the editor who the receptionist who quit married admired.

These sentences all require the NP *the book* to be assigned as the object of the final verb *admired*, differing only on the amount of distance between the filler and the role assigner. They found that while overall response accuracy was affected by the amount of interpolated material, when subjects were able to connect the filler with its role assigner, the time to do so was constant (see also (McElree, Foraker, & Dyer, in press)). In other words, the intervening material may affect whether or not an item is available in memory, but when it is, it will be accessed directly and not as the result of a sequential search process. Hence, the Vosse and Kempen parser appears to be correct at least regarding direct access to potential fillers. However, since intervening material may influence the amount of link competition, and hence the time to resolve alternative links, it is unclear whether the model accurately captures the McElree data. In contrast, the parsing as retrieval approach accounts for these results easily, since the principle of global matching entails a content-addressable search in which cues are combined equally and compared simultaneously with the content of all items in memory. Since the probability with which an item will be accessed depends not only on the grammatical and semantic retrieval cues, but also on the availability of an item in working memory (i.e., its self-strength), McElree's finding of distance effects on accuracy is predicted by this approach as well.

4.2 Ambiguity resolution

Since the parsing as retrieval approach and the Unification Space parser both incorporate parallelism into the process of identifying potential links, competition between competing links at points of ambiguity is inevitable. As mentioned above, competition in the Unification Space parser determines whether a particular link wins out over alternatives, suggesting that certain links may be easier to create than others. In a parser based on retrieval, competition comes from the quality of the cues used in the retrieval, and whether or not they can unambiguously identify the proper item to be linked. If cues are too ambiguous, then the probability of selecting the correct item is reduced and the correct link may not be proposed at all. This will not occur in the Vosse and Kempen model, since all links are proposed, although competition from stronger links may prevent a particular link from being realized.

This distinction is important in an account of why certain structures can be reanalyzed more easily than others. For example, Sturt, Pickering, and Crocker (1999) compared sentences such as those in Example 4.4 with sentences like in Example 4.5, which require a different sort of reanalysis to correct the initial misinterpretation of *doctor* as the object of the verb *saw.* In the following discussion we refer to sentences like Example 4.4 as NP/Z sentences because the ambiguous verb *visited* can take either an NP complement or no (Zero) complement. Similarly, sentences like Example 4.5 will be referred to as NP/S sentences, since the verb *saw* can take either an NP or an Sentential complement.

Example 4.4: Before the woman visited the doctor had been drinking quite a lot.

Example 4.5: The woman saw the doctor had been drinking quite a lot.

In both these examples, the verb *had* initiates the creation of a sentential structure (i.e., an IP projection), which must be incorporated into the existing parse tree. Hence, it provides retrieval cues for an item looking for an IP projection. In Example 4.5 this is satisfied by the reading of the verb *saw* which was not chosen in the initial interpretation, and although this alternative lexical frame has begun to decay, if the distance between *saw* and *had* is not too long, as in this example, then it is retrieved and *had* can be integrated into the sentence (cf. (Van Dyke & Lewis, submitted) for data on distance effects on reanalysis). At this point, the existing grammatical dependencies are as represented in Figure 4.2.

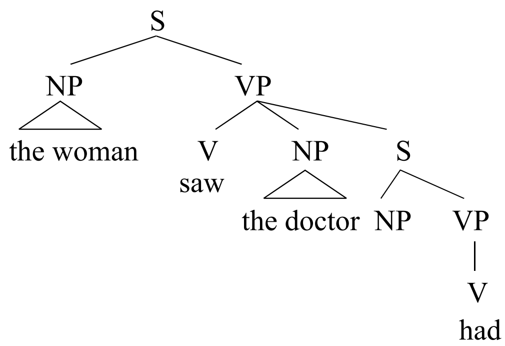


Figure 4.2 NP/S reanalysis before detaching the incorrect link

Although *had* has been integrated into the current parse, its grammatical roles are still unsatisfied, and so it provides retrieval cues for a subject NP, which is satisfied by the NP *the doctor*, once it is detached from the now less active lexical frame. This allows the reanalysis to be completed and the rest of the sentence can be interpreted fairly easily, as indicated by faster reading times (Sturt et al., 1999) and lower error rates (Van Dyke & Lewis, submitted) as compared with NP/Z sentences like Example 4.4.

The difficulty of reanalyzing NP/Z sentences can be explained by insufficient retrieval cues (see Van Dyke and Lewis (submitted) for further discussion). When the verb *had* occurs in this example, it can be integrated into the sentence because the adjunct clause *Before the woman visited the doctor* expects to modify a main clause, as in Figure 4.3.

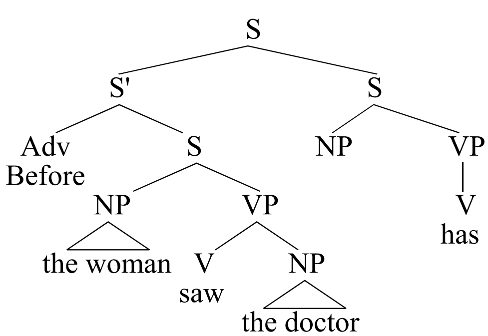


Figure 4.3: NP/Z reanalysis before detaching the incorrect link.

After it has been integrated, the verb *has* still must find a subject, so it supplies retrieval cues for the NP *the doctor*, which is attached as its subject, resulting in the structure illustrated in Figure 4.4.

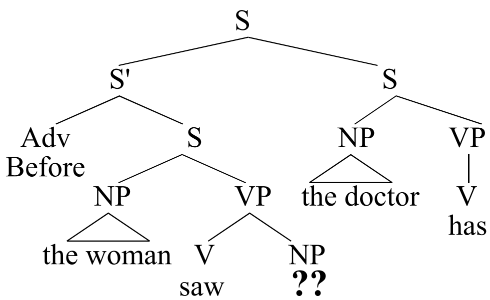


Figure 4.4: Final state of the NP/Z reanalysis.

Critically, this structure is not a well-formed parse tree because the expected object of the verb *saw* is unfulfilled. The difficulty with this structure has been described by Fodor and Inoue (1998) as the Thematic Overlay Effect, in which the parser prefers not to demote a semantic argument role from overt (lexically realized) to implicit (unrealized). The only way this structure could be made acceptable is if the alternative, non-transitive, lexical frame of the verb *saw* were substituted for the transitive frame, but in this case, there are no retrieval cues for doing so. This makes this reanalysis much more difficult than the NP/S reanalysis, in which the verb *has* provides all the necessary retrieval cues.

In a model such as the Vosse and Kempen model, which does not account for insufficient retrieval cues, the explanation for this contrast is ad hoc and difficult to interpret. They suggest that the relative difficulty of reanalyzing NP/Z structures is because the incorrect attachment of the NP *the doctor* as the object of the verb *saw* easily wins out over the competing link of *the doctor* as the subject of the main clause. Vosse and Kempen report that this occurs when the verb is *mildly* biased towards taking an NP complement, meaning that they have set the activation strengths of the links in the model to reflect this bias. This is critical because the activation strengths of alternative links determine the amount of competition between them, which in turn determines how easily the model settles on the correct analysis.

In particular, a mild bias means that the competition will be quite fierce, since there is no clear bias for one or the other structures. Moreover, it is unclear whether the difficulty observed in the Vosse and Kempen model is due to reanalysis or simply to difficulty settling on the correct initial interpretation. Sturt et al. (1999) showed that NP/Z reanalyses are more difficult than NP/S reanalyses when the mean NP bias was 89%, which is not at all a "mild" bias. From this it would appear that the parser has a very strong commitment to a direct object interpretation of the NP *the doctor,* making reanalysis necessary when it turns out to be wrong. Since Vosse and Kempen do not provide a reference point for the bias used in their model as compared to actual distributional statistics, it is unclear whether the model accurately accounts for the Sturt et al. reanalysis effects.

4.3 Nature of the Lexicon

Following other constraint-based approaches (e.g., (MacDonald et al., 1994; Trueswell et al., 1994; Trueswell & Tannenhaus, 1994), the Vosse and Kempen parser assumes a fully lexicalized grammar, meaning that lexical entries are actually phrase structure trees, projected to their highest maximal projection (cf. Figure 4.1). This type of lexicon is what allows the unification mechanism to succeed, and eliminates the need for a separate parsing mechanism that applies phrase structure rules to create linguistic structure. Although this may sound appealing as a parsimonious description of the language processor, it actually necessitates an extreme amount of redundancy in the lexicon. For example, the lexical entry for *sees* in Figure 4.1 specifies a PP modifier, which allows it to account for the sentence *The woman saw the man with the binoculars*. Yet, a different kind of modifier would require an entirely different lexical entry, as in the sentence *The woman saw the man only briefly* or *The woman saw the man smiling.* In each case, unification of the modifier with the verb *see* can occur only if the modifier is specified with the proper grammatical category (i.e. adverb or adjective phrase).

Evidence for this level of lexicalization has been difficult to find. Some tentative evidence appeared in a study by Spivey-Knowlton and Sedivy (1995), who argued for the presence of *with*-PP's in the lexical entries for action verbs based on the comparison of Example 4.6 and Example 4.7.

Example 4.6: The mechanic changed a tire with a faulty valve.

Example 4.7: The mechanic changed a tire with a monkey wrench.

Reading times for the prepositional phrase were more difficult in Example 4.6, in which it modifies the object NP than in Example 4.7. In contrast, the opposite pattern was found in sentences containing perception verbs, as in Example 4.8 and Example 4.9.

Example 4.8: The salesman glanced at a customer with ripped jeans.

Example 4.9: The salesman glanced at a customer with suspicion.

They concluded that action verbs are lexically specified for a *with*-PP modifier, which causes a garden path effect when the *with*-PP modifies the NP instead of the verb in Example 4.8. In contrast, perception verbs are not specified for a *with*-PP, and when they are modified with such a PP, as in Example 4.9, processing is more difficult because there is no readily available attachment site in the lexical entry. This conclusion has been questioned on a number of grounds (cf. (Blodgett & Boland, in press; Boland & Boehm-Jenigan, 1998), but the most direct challenge has come from Boland, Lewis, Sudhakar, and Blodget (in preparation). They argued that Spivey-Knowlton and Sedivy's conclusion was unjustified because the critical verb-type x attachment site interaction could not be tested since the different verb types were tested in separate experiments. Moreover, VP attachment with action verbs is pragmatically more likely, since actions are often done with instruments. Hence the relative ease of processing the PP when it is attached to the verb is not because the modifier is encoded in the lexicon, but simply because that's where it is more likely to be attached. To address these concerns, Boland et al. conducted an experiment that manipulated both verb bias and attachment site in the same experiment, plus conditions controlling for pragmatic effects on attachment, and for the possibility that the *with*-PP is actually an argument and not an adjunct. They found no data supporting the conclusion that PP modifiers are represented in the lexical entries of either action or perception verbs. Rather, their results supported the encoding of argument structure *only* in the lexicon (cf. Lewis and Boland (under revision) for further discussion), suggesting that a mechanism in addition to simple unification (i.e., a parser) is necessary to account for grammatical structure building, especially in the case of modifiers.

In contrast to the Vosse and Kempen parser, and fitting with the Boland et al. results, the cue-based retrieval parser does not require a fully grammaticized lexicon. Rather, it requires only the specification of argument structure in order to provide retrieval cues for the correct number and type of grammatical dependencies. In addition, since retrieval cues are assembled from semantic, pragmatic, and grammatical features simultaneously, lexical entries for this parser are more akin to the SYNSEM multi-leveled feature structures assumed by head-driven phrase structure grammar (Pollard & Sag, 1994). With such a representation, the retrieval cues specified by a grammatical head can contain a description of the desired filler with both grammatical and semantic properties.

One particularly intriguing consequence of this is that the cue-based retrieval parser can provide a unified processing account of the argument/adjunct asymmetry regarding the influence of pragmatic information on initial attachment. For example, Britt et al. (1992) investigated the extent to which discourse context influences attachments of ambiguous adjunct prepositions such as those in Example 4.10, as compared to ambiguous argument attachments in sentences such as Example 4.11.

Example 4.10: The doctor examined the child with the needle.

Example 4.11: The editor played the tape agreed the story was big. (Ferreira & Clifton, 1986)

They found that the attachment of the adjunct preposition was influenced by context, however, this was not the case for argument attachments (see also (Ferreira & Clifton, 1986)). Results such as these have led Frazier and Clifton (1996) to propose a bifurcated processing model in which arguments are *attached* according to strictly grammatical principles, while adjuncts are *associated* by a construal mechanism that uses both grammatical and semantic/pragmatic information.

With a cue-based retrieval parser, this distinction falls out naturally when considering the nature of retrieval cues for the two types of attachment. In the case of arguments, the argument structure of the grammatical head specifies cues for particular grammatical roles (e.g., subject, object, etc.) and these cues will dominate the search for the correct dependent. Modifiers, on the other hand, augment properties of a particular host and will therefore initiate a search for that host, but since the grammatical specification of the type of host may be ambiguously defined, semantic or pragmatic cues will play a greater role in selecting among candidates. Thus, unlike a parser working from a lexicalized grammar (e.g., Vosse and Kempen), the cue-based retrieval parser makes it is unnecessary to specify modifiers in the lexical entry of each potential host because they will be identified using the semantic cues coming from the adjunct itself.

4.4 Does the parser parse?

The Vosse and Kempen model joins other constraint-based approaches in attempting to do away with a parser that is specialized for building grammatical structure. This occurs in two ways: first, putting grammatical knowledge in the lexicon means the structure builder can be dumb with respect to this knowledge, and secondly, no actual constituent structure need be built because matching feature structures are simply unified. The cue-based retrieval parser provides an interesting compromise between these approaches, resulting in quite a different view of the job of the parser. On the one hand, we have seen that detailed grammatical knowledge is not contained in the lexicon of the retrieval parser, suggesting that some grammar-specific processor is necessary. Yet, on the other, this processor is not required for building structure between grammatical dependents, as this job is done by the cue-based retrievals initiated by grammatical heads. This means that the retrieval parser does not parse, when parsing refers to structure building. Rather, the job of the parser is simply as an interpreter of linguistic context. For example, the lexical entry of a verb will specify syntactic retrieval cues for a subject, but these cues are only useful when the candidate items are annotated for subjecthood. The level of knowledge required for making these annotations is much more shallow than that usually attributed to the parsing mechanism, and can be described simply as the ability to assign grammatical Case (Chomsky, 1981; Haegeman, 1991).

Evidence illustrating the import of a separate Case assigning mechanism to accompany cue-based parsing is seen in a study by Trueswell, Tanenhaus, and Kello (1993). They used sentential complement verbs that typically appear either with an explicit complementizer (Example 4.12) or without a complementizer (Example 4.13).

Example 4.12: The man wished (that) the award would go to his brother.

Example 4.13: The man hinted (that) the award would go to his brother.

Comparing reading times for the NP *the award* when the complementizer was present or not, they found a complementizer effect that was strongly correlated with the frequency with which the verb appeared with or without the complement. Thus, when the verb provided strong cues that the NP *the award* would be subject of its clausal complement, the NP was read more quickly, even when the complementizer was not overt. This is especially important from the point of view of cue-based retrieval, because prior to the occurrence of *would*, there has been no retrieval that would attach *the award* into the parse tree. Specifically, the retrieval cues for the matrix verb are expecting a clause, and not an NP, and hence will not match to *award* before it has been joined with its verb. Rather than supposing that this NP remains unprocessed until its verb occurs, a situation that is contrary to the empirical results described here, we suggest that the parsing mechanism is able to assign grammatical Case to the NP on the bases of cues from the matrix verb even though it has not been fully integrated into the parse tree. This in turn facilitates its later retrieval as the subject of *would*, since it will now match with the argument structure requirements of that verb. In the case when the matrix verb does not clearly support assignment of Case, reading times are much slower on the NP because the parser doesn't have enough contextual information to assign case, and this in turn makes reading times at *would* slower because the NP does not make a clear match for its subject role. Thus, although the specification of argument structure in the grammatical head is sufficient to initiate retrievals for its required dependents, the extent to which those retrievals will succeed depends critically on whether the linguistic context allows the parser to assign the grammatical Case the retrieval cues are looking for.

4.5 Retrieval parsing and interpretation

A key assumption of cue-based parsing is that each word in a sentence is represented as a feature structure, built up from the semantics of the item itself, and from elaboration from subsequent predication of both semantic and syntactic properties (i.e., grammatical case). This provides the retrieval context for interpreting the cues available from either the anaphor or the grammatical head initiating the retrieval. Figure 4.5 illustrates how this might work, with each memory item being compared simultaneously, according to Equation 4.1, with the cues relevant for the retrieval probe (indicated in bold in Figure 4.4).

$$P(I_i|Q_1, \ldots, Q_m) = \frac{\prod_{j=1}^{m} S(Q_j, I_j)^{w_j}}{\sum_{k=1}^{N} \prod_{j=1}^{m} S(Q_j, I_k)^{w_j}}$$

Equation 4.1: SAM retrieval function

The item having the highest match with the features of the retrieval probe will also have the highest probability of retrieval, and it is this item which will complete the grammatical dependency headed by *asleep*. Retrieval probabilities for each item in Example 4.14 are given in the last row of Figure 4.5.

Example 4.14: The girl saw that the boy who was sitting in the comfortable seat was asleep.

Figure 4.5: Working memory feature structure for interpreting retrieval cues from *asleep*.

Indexed Cues		Memory Item						
		NP_1	V1	NP2	V2	P	NP3	V4
Q_1	**$\text{Syntax}_{\text{sNoun}}$**	.99	.01	.95	.01	.01	.98	.01
Q_2	$\text{Syntax}_{\text{Verb}}$	.01	.67	.01	.98	.01	.01	.97
Q_3	$\text{Syntax}_{\text{Prep}}$	.01	.01	.01	.01	.99	.01	.01
Q_4	**$\text{Syntax}_{\text{Subject}}$**	.92	.03	.92	.02	.02	.02	.02
Q_5	$\text{Syntax}_{\text{V-object}}$	.02	.02	.23	.82	.01	.02	.01
Q_6	$\text{Syntax}_{\text{Prep-object}}$	.01	.02	.03	.01	.02	.97	.03
Q_7	$\text{Semantics}_{\text{saw}}$	.78	.99	.82	.02	.01	.06	.02
Q_8	$\text{Semantics}_{\text{sitting}}$	[illegible]	.04	.07	.99	.45	.04	.02
Q_9	$\text{Semantics}_{\text{next-to}}$	.04	.02	.12	.67	.98	.02	.01
Q_{10}	**$\text{Semantics}_{\text{asleep}}$**	.96	[illegible]	.95	.02	.02	.02	.99
Q_n	Other cues	.76	.34	.13	.02	.11	.70	.65
Q_{n+1}	**$\text{Self-strength}_{\text{recency}}$**	.30	.40	.50	.60	.70	.80	.90
Retrieval probability (asleep)		.26	.00	.42	[illegible]	.00	.00	.00

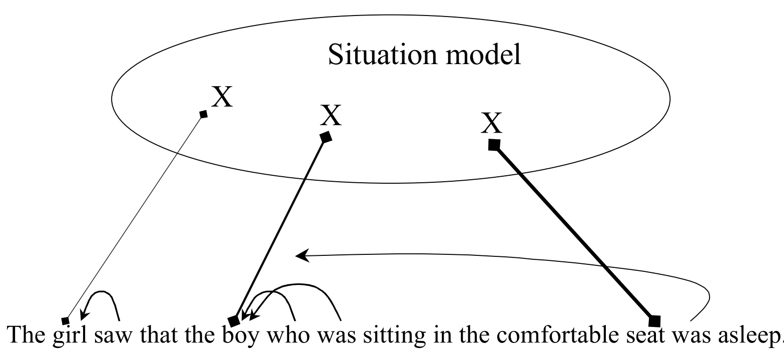
[**asleep**:
+animate,
+subject,
+N,
+sleepable]

In this example, there is no difficulty retrieving the correct subject of the verbal *was asleep* because there are no syntactically or semantically interfering items. However, the creation of this dependency is merely the first step in comprehending the sentence, as the new dependency must now be integrated into the discourse context. Thus, in Example 4.14, when *the boy* is selected as the subject of *was asleep*, this new clause consists of a meaning proposition which must be augmented onto the representation of what *the girl saw*.

While creating the dependency was easy because global matching identified the correct grammatical subject, elaboration of the situation model is difficult because the discourse referent for *saw* has decayed and no intermediate retrievals have occurred which would reactivate it in memory. Moreover, elaboration at the level of the situation model requires more than simply completing the grammatical dependency between *saw* and the clausal complement, as additional inferences such as the likelihood of the comfortable seat being the reason that the boy fell asleep may also be created. Thus, there are two distinct operations which must occur in order to create comprehension: completion of the grammatical dependency and elaboration of the situation model. These two operations are analogous to the distinction between *bonding* and *resolution* of anaphors, proposed by Garrod and colleagues (Garrod, Freudenthal, & Boyle, 1994; Garrod & Terras, 2000; Sanford & Garrod, 1989). They proposed that bonding is a process of creating the initial link between an anaphor and its antecedent, while resolution is the process of interpreting the bond in the discourse context.

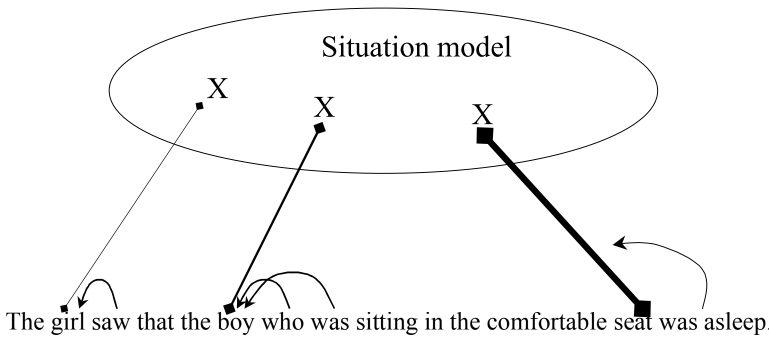
The data in Experiment 5 suggests an interesting interplay between these two processes. Specifically, when the retrieval cues from the verb are unambiguous, and the intervening NP is less active, retrieval interference is minimized and the correct NP can be selected out of the retrieval matrix illustrated in Figure 4.5. However, in order to elaborate this NP with the semantics of its new predicate, it must be activated, and this is costly because it has a comparatively low self-strength, as illustrated in Figure 4.6.

Figure 4.6: Activation levels when *seat* is *given*, making its self-strength low.



In contrast, when the intervening NP is highly activated, it causes retrieval interference that makes selecting the correct NP from the retrieval matrix more difficult, as in Figure 4.7.

Figure 4.7: Activation levels when *seat* is *new*, making its self-strength high.



Since the incorrect NP is selected for the grammatical dependency, elaboration of this NP with the new predicate is easy because it is already highly active. The result is an incorrect situation model that was comparatively easy to construct, explaining the reduced reading times in the corresponding region.

Thus, both parsing and elaboration of the situation model are dependent on easy retrieval for their success. This raises the possibility of a unified account of both parsing and interpretation in terms of the memory mechanism that must support comprehension. Moreover, it raises the possibility that separate syntactic-level representations may be unnecessary, since the retrieval mechanism can access items from the discourse context to create grammatical dependencies, which are then integrated more fully into the situation model. While further data is required to evaluate these hypotheses, the data provided in the 5 experiments here present preliminary evidence in support of the parsing as retrieval proposal, suggesting that implications of this approach warrant further consideration as well.

4.6 Storage effects (revisited)

Notably, the experiments in this dissertation provided no evidence for storage effects in sentence processing. I do not mean to suggest that these effects are not present, as it is reasonable to assume some measurable effects due to storing more information rather than less could be found. However, I am suggesting that these effects are not the *source* of processing breakdown, but rather it is the efficiency with which available retrieval cues enable creation of grammatical dependencies. This conclusion fits with the data presented here, as well as the data reviewed in Chapter 1 suggesting that complexity effects occur at retrieval points rather than storage points. In future work, it will be interesting to see whether further investigations can tease apart retrieval effects from storage effects, especially using more subtle neurophysiological dependent measures.

APPENDICES

Appendix A

Experiment 2: Raw data and analyses

Table A.1: Chapter 2, Section 2.4 raw reading times, analysis with subjects as the random factor.

		Region 1	Region 2	Region 3	Region 4
No distance manipulation Interference Type					
	Low syntactic, Low semantic	2154 (103)	3044 (137)	924 (38)	566 (30)
	Low syntactic, High semantic	2125 (101)	3044 (157)	963 (38)	583 (24)
	High syntactic, Low semantic	2175 (88)	3408 (161)	960 (41)	618 (35)
	High syntactic, High semantic	2130 (107)	3335 (184)	1060 (52)	575 (30)
With distance manipulation Interference Type					
	Low syntactic, Low semantic	2163 (110)	4306 (215)	858 (42)	597 (24)
	Low syntactic, High semantic	2069 (99)	4417 (234)	921 (44)	559 (28)
	High syntactic, Low semantic	2144 (100)	4419 (217)	888 (36)	643 (26)
	High syntactic, High semantic	2079 (108)	4492 (256)	1072 (71)	689 (43)

Table A.2: Chapter 2, Section 2.4 raw reading times, analysis with items as the random factor.

		Region 1	Region 2	Region 3	Region 4
No distance manipulation Interference Type					
	Low syntactic, Low semantic	2115 (81)	2983 (132)	915 (48)	547 (37)
	Low syntactic, High semantic	2148 (96)	3037 (143)	943 (40)	560 (41)
	High syntactic, Low semantic	2106 (85)	3327 (139)	934 (46)	583 (47)
	High syntactic, High semantic	2082 (89)	3284 (155)	1038 (58)	564 (41)
With distance manipulation Interference Type					
	Low syntactic, Low semantic	2121 (88)	4255 (184)	863 (35)	585 (41)
	Low syntactic, High semantic	2092 (91)	4211 (158)	888 (39)	578 (42)
	High syntactic, Low semantic	2012 (85)	4314 (238)	879 (36)	607 (42)
	High syntactic, High semantic	2170 (158)	4769 (284)	1144 (103)	671 (49)

Statistical analyses

No significant effects were found in Region 1. In Region 2, a main effect of syntactic interference was found, $F_1(1,44) = 6.27$, $p < .02$; $F_2(1,30) = 5.34$, $p < .03$. This effect is likely due to the fact that the low syntactic interference sentences had 5 words and the high syntactic interference sentences had 6 words (or 8 and 9 words with the distance manipulation). The residual analyses reported in Section 2.4 corrected for this confound. Region 2 also exhibited an effect of distance, $F_1(1,44) = 119.17$, $p < .001$; $F_2(1,30) = 89.60$, $p < .001$. This is because region 2 is locus of the distance manipulation. The critical region for retrieval of the long distance subject is Region 3, where the main effect of syntactic interference was observed, $F_1(1,44) = 9.64$, $p < .004$; $F_2(1,30) = 5.19$, $p < .04$. A main effect of semantic interference was also observed $F_1(1,44) = 22.14$, $p < .001$; $F_2(1,30) = 11.23$, $p < .003$. The interaction of syntactic x semantic interference was also significant, $F_1(1,44) = 4.94$, $p < .04$; $F_2(1,30) = 4.38$, $p < .05$. In Region 4, the wrap-up region, a main effect of syntactic interference was observed, $F_1(1,44) = 10.67$, $p < .003$; F_2 $(1,30) = 5.34$, $p < .03$. An effect of distance was also observed in this region in the items analysis, $F_2(1,30) = 9.87$, $p < .005$, but not in the subject analyses $F_1(1,44) = 3.14$, $p < .09$. However, the subject analyses did show two interactions with the distance variable that were not present in the items analyses. The syntactic interference x distance interaction was significant by subjects, $F_1(1,44) = 4.57$, $p < .04$; $F_2(1,30) = 1.79$, $p < .20$. In addition, the three-way interaction of syntactic and semantic interference and distance was significant in the subject analyses, $F_1(1,44) = 4.97$, $p < .04$, but not for the items analysis, $F_2(1,30) = 2.31$, $p < .14$. These two interactions are due to the large effect of distance in the presence of syntactic interference ($M_1 = 596$ with no distance manipulation, $M_1 = 666$ with the manipulation), but not when syntactic interference was low ($M_1 = 574$ with no distance manipulation, and $M_1 = 568$ with the manipulation). This effect would be reduced when regression analyses correct the confound of region length and the syntactic manipulations in region 2, thus explaining not only this discrepancy with the results on residuals reported in Chapter 2 but also the effect of distance in region 2 which was found here, but not in the residual analyses. With the exception of these effects, the pattern of significant effects observed here is the same as that reported in Chapter 2.

Appendix B

Experiment 5: Raw data and analyses

Table B.1: Trimmed residual reading times for each region (ms), subjects as random factor.

	Region 1	Region 2	Region 3	Region 4
Low syntactic interference				
New	1925 (87)	2434 (111)	807 (38)	543 (26)
Given	1837 (68)	2253 (91)	745 (34)	615 (27)
Indexical	1876 (83)	1985 (98)	773 (32)	542 (28)
High syntactic interference				
New	1889 (90)	2820 (143)	845 (46)	571 (28)
Given	1863 (79)	2638 (107)	801 (24)	593 (22)
Indexical	1892 (83)	2548 (125)	899 (51)	562 (26)

Table B.2: Trimmed residual reading times for each region (ms), items as random factor.

	Region 1	Region 2	Region 3	Region 4
Low syntactic interference				
New	1932 (51)	2422 (76)	799 (28)	545 (28)
Given	1848 (52)	2275 (56)	748 (27)	612 (31)
Indexical	1890 (55)	2010 (64)	783 (24)	547 (30)
High syntactic interference				
New	1887 (49)	2823 (79)	847 (24)	567 (28)
Given	1876 (52)	2646 (66)	806 (25)	591 (28)
Indexical	1890 (55)	2570 (89)	900 (33)	556 (30)

Statistical Analysis

No significant effects were observed in Region 1. In Region 2, the effect of referent type was significant, $F_1(2,70) = 33.88$, $p < .001$; $F_2(2,98) = 25.38$, $p < .001$, mirroring the residual analysis reported in Chapter 4. It should be noted, however, that this effect is confounded with number of words since the indexical conditions had one fewer word than the others. In addition, the difference in the linear position of the new and indexical conditions as compared with the given condition, which occurs after more words because of the initial clause, may have contributed to shorter reading times in the given condition. The regression analysis described in Chapter 4 corrected for these confounds. The effect of interference type was significant in this region, $F_1(1,35) = 152.31$, $p < .001$; $F_2\ (1,49) = 95.60$, $p < .001$. Again, this is likely due to the fact that the high interference sentences have one extra word as compared with the low interference sentences. Moreover, as in Chapter 5, the difference between the low and high interference sentences in the indexical conditions was much greater than for the other two conditions, supporting the suggestion proposed in Chapter 5 that processing the indexical was more difficult when it occurred in a high interference sentence. This difference led to a significant interaction of referent type and interference type for the subject analyses, $F_1(2,70) = 3.17$, $p < .05$, but not for the item analysis, $F_2(2,98) = 1.82$, $p = .17$.

In region 3, the effect of referent type was significant, $F_1(2, 70) = 5.14$, $p < .009$; $F_2(2,98) = 9.07$, $p < .001$, due to the given condition being read more quickly. The effect of interference type was also significant, again due to the high interference sentence being more difficult, $F_1(2,70) = 13.98$, $p < .001$; $F_2(1,49) = 23.33$, $p < .001$. There was a trend for a significant interaction of referent type and interference type in the subject analysis, $F_1(2,70) = 2.60$, $p =.08$, but not by items, $F_2(2,98) = 1.91$, $p =.15$. These results closely mirrored the residual results reported in Chapter 5. The difference between the given and new conditions was significant for the low interference conditions, but not for the high interference conditions in the subject analysis; however, the difference between the given and new conditions in the items analysis did not quite reach significance ($p =.10$).

In region 4, only the effect of referent type was significant, $F_1(2,70) = 4.45$, $p < .02$; $F_2(2,98) = 10.95$, $p < .001$, due to the given condition being read more slowly. As with the residual analyses, the interaction of region type and referent type was significant, $F_1(2,70) = 13.62$, $p < .001$; $F_2(2,98) = 26.70$, p

< .001, due to the condition that was easy in Region 3 become more difficult in Region 4, and vice versa. Table 35 presents the raw means for this analysis. Thus, with the exception of a few effects made stronger because of length or position differences in the materials, the effects observed in the raw reading time analyses are the same as those reported in Chapter 5 using residuals.

Table B.3: Experiment 5, means for region x referent type interaction.

	Means with subjects as random factor		Means with items as random factor	
	Region 3	Region 4	Region 3	Region 4
New	826 (40)	557 (25)	823 (22)	556 (27)
Old	774 (26)	604 (22)	777 (23)	602 (28)
Indexical	836 (39)	552 (25)	842 (25)	552 (29)

BIBLIOGRAPHY

Bibliography

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

Almor, A. (1999). Noun-phrase anaphor and focus: The informational load hypothesis. Psychological Review, 106(4), 748-765.

Anderson, J. R., & Lebiere, C. (1998). The Atomic Components of Thought. Mahwah, NJ: Erlbaum.

Ariel, M. (1990). Accessing Noun-Phrase Antecedents. London: Routledge.

Ariel, M. (1994). Interpreting anaphoric expression: A cognitive versus a pragmatic approach. Journal of Linguistics, 30, 3-42.

Baddeley, A. (1992). Working memory. Science, 255, 556-559.

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic, and formal similarity. Quarterly Journal of Experimental Psychology, 18, 362-365.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), The psychology of learning and motivations, vol. 8: Advances in research and theory : Academic Press.

Bader, M., & Meng, M. (1999). Subject-object ambiguities in German embedded clauses: An across-the-board comparison. Journal of Psycholinguistic Research, 28(2), 121-143.

Bever, T. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), Cognition and the Development of Language . New York: Wiley.

Blodgett, A., & Boland, J. E. (in press). Argument status and attachment site in prepositional phrase attachment. .

Boland, J. E., & Boehm-Jenigan, H. (1998). Lexical constraints and prepositional phrase attachment. Journal of Memory and Language, 39(4), 684-719.

Boland, J. E., Lewis, R. L., Sudhakar, M., & Blodgett, A. (in preparation). Lexical parsing: Limits on lexicalization of syntactic generation. .

Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. Journal of Memory and Language, 34, 774-806.

Booth, J. R., MacWhinney, B., & Harasaki, Y. (2000). Developmental differences in visual and auditory processing of complex sentences. Child Development, 71(4), 981-1003.

Bower, G. H., & Rinck, M. (2001). Selecting one among many referents in spatial situation models. Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(1), 81-98.

Bowles, N. L., & Glanzer, M. (1983). An analysis of interference in recognition memory. Memory & Cognition, 11(3), 307-315.

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

Caplan, D. (1985). Comments on Fodor's The modularity of Mind: A neo-Cartesian alternative. Behavior and Brain Sciences, 8, 6-7.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. Behavioral and Brain Sciences, 22, 77-126.

Carpenter, P. A., Miyake, A., & Just, M. A. (1994). Working memory constraints in comprehension. In M. A. Gernsbacher (Ed.), Handbook of Psycholinguistics (pp. 1075-1122). New York: Academic Press.

Chafe, W. (1987). Cognitive constraints on information flow. In R. Tomlin (Ed.), Coherence and grounding in discourse (pp. 21-51). Philadelphia, PA: John Benjamins.

Chomsky, N. (1981). Lectures on Government and Binding. Dordrecht: Foris.

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.

Clark, H., & Sengul, C. J. (1979). In search of referents for nouns and pronouns. Memory & Cognition, 7(1), 35-41.

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.

Conrad, R. (1963). Acoustic confusions and memory span for words. Nature, 197, 1029-1030.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior, 19, 450-466.

Dell, G. S., McKoon, G., & Ratcliff, R. (1983). The activation of antecedent information during the processing of anaphoric reference in reading. Journal of Verbal Learning and Verbal Behavior, 22, 121-132.

Enç, M. (1983). Anchored expressions. Paper presented at the West coast conference of formal linguistics.

Faneslow, G., Kliegl, R., & Schlesewsky, M. (1999). Processing difficulty and principles of grammar. In S. Kemper & R. Kliegl (Eds.), Constraints on language (pp. 171-201). Dordrecht: Kluwer Academic.

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

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

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2001). Syntactic working memory and the establishment of filler-gap dependencies: Insights from ERPs and fMRI. Journal of Psycholinguistic Research, 30(3), 321-338.

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (submitted). An ERP investigation of syntactic working memory during the processing of German WH-questions. .

Fodor, J. A. (1983). The Modularity of Mind. Cambridge, MA: MIT Press.

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. (1983). A method for obtaining measures of local parsing complexity throughout sentences. Journal of Verbal Learning and Verbal Behavior, 22, 203-218.

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

Frazier, L., & Clifton, C. (1996). Construal. Cambridge, MA: MIT Press.

Frazier, L., Clifton, C., & Randall, J. (1983). Filling gaps: Decision principles and structure in sentence comprehension. Cognition, 13(2), 187-222.

Garrod, S., Freudenthal, D., & Boyle, E. (1994). The role of different types of anaphor in the on-line resolution of sentences in a discourse. Journal of Memory and Language, 33, 39-68.

Garrod, S., & Sanford, A. J. (1977). Interpreting anaphoric relations: The integration of semantic information while reading. Journal of Verbal Learning and Verbal Behavior, 16, 77-90.

Garrod, S., & Sanford, A. J. (1982). Bridging inferences in the extended domain of reference. In A. Baddeley & J. Long (Eds.), Attention and performance. IX. (Vol. IX, pp. 331-346). Hillsdale, NJ: Erlbaum.

Garrod, S., & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and Resolution. Journal of Memory and Language, 42, 526-544.

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. Cognition, 32, 99-156.

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 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., & 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., & Johnson, M. (in press). Memory interference during language processing. Journal of Experimental Psychology: Learning, Memory, and Cognition.

Gorrell, P. (1993). Evaluating the direct association hypothesis: A reply to Pickering and Barry (1991). Language and Cognitive Processes, 8(2), 129-146.

Graf, P., & Mandler, G. (1984). Activation makes words more accessible, but not necessarily more retrievable. Journal of Learning and Verbal Behavior, 23, 553-568.

Gundel, J. K., Hedberg, N., & Zacharaski, R. (1993). Cognitive status and form of referring expressions. Language, 69(2), 274-307.

Haberlandt, K., & Graesser, A. C. (1989). Processing of new arguments at clause boundaries. Memory & Cognition, 17(2), 186-193.

Haegeman, L. (1991). Introduction to Government & Binding Theory. Cambridge, MA: Blackwell.

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. Journal of Verbal Learning and Verbal Behavior, 13, 512-521.

Hawkins, J. A. (1999). Processing complexity and filler-gap dependencies across grammars. Language, 75(2), 244-285.

Hickok, G. (1993). Parallel parsing: Evidence from reactivation in garden-path sentences. Journal of Psycholinguistic Research, 22, 239-250.

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

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

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

Humphreys, M. S. (1976). Relational information and the context effect in recognition memory. Memory and Cognition, 4, 221-232.

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

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. Psychological Review, 99(1), 122-149.

Just, M. A., Carpenter, P. A., & Keller, T. A. (1996a). The capacity theory of comprehension: New frontiers of evidence and arguments. Psychological Review, 103(4), 773-780.

Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996b). Brain activation modulated by sentence comprehension. Science, 274, 114-116.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. Language and Cognitive Processes, 15(2), 159-201.

Kamide, Y., & Mitchell, D. C. (1999). Incremental pre-head attachment in Japanese parsing. Language and Cognitive Processes, 14, 631-662.

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.

King, J., & Kutas, M. (1995). Who did what and when? Using word and clause related ERPs to monitor working memory usage in reading. Journal of Cognitive Neuroscience, 7, 378-397.

Kintsch, W. (2001). Predication. Cognitive Science, 25(2), 173-202.

Lewis, R., & Boland, J. E. (under revision). In search of fully lexical parsing. .

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.

Logie, R. H., Zucco, G. M., & Baddeley, A. D. (1990). Interference with visual short-term memory. Acta Psychologica, 75, 55-74.

Love, T., & Swinney, D. (1996). Coreference processing and levels of analysis in object-relative constructions; Demonstration of antecedent reactivation with the cross-modal priming paradigm. Journal of Psycholinguistic Research, 25(1), 5-24.

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

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

MacWhinney, B., & Pleh, C. (1988). The processing of restrictive relative clauses in Hungarian. Cognition, 29, 95-141.

Martin, R. C., & Feher, E. (1990). The consequences of reduced memory span for comprehension of semantic vs. syntactic information. Brain and Language, 38(1-20).

Martin, R. C., & Freedman, M. L. (in press). Short-term retention of lexical-semantic representations: Implications for speech production. Memory.

Martin, R. C., Lesch, M. F., & Bartha, M. C. (1999). Independence of input and output phonology in word processing and short-term memory. Journal of Memory and Language, 40, 1-27.

Martin, R. C., & Romani, C. (1994). Verbal working memory and sentence comprehension: A multiple-components view. Neuropsychology, 8(4), 506-523.

Martin, R. C., Shelton, J. R., & Yaffee, L. S. (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. Journal of Memory and Language, 33(83-111).

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. (in press). Memory structures that subserve sentence comprehension. .

McKoon, G., & Ratcliff, R. (1980). The comprehension processes and memory structures involved in anaphoric reference. Journal of Verbal Learning and Verbal Behavior, 19, 668-682.

Mensink, G., & Raaijmakers, J. (1988). A model of interference and forgetting. Psychological Review, 95(4), 434-455.

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: John Wiley.

Murdock, B. B., Jr. (1982). A theory for the storage and retrieval of item and associative information. Psychological Review, 89, 609-626.

Murdock, B. B., Jr. (1983). A distributed memory model for serial-order information. Psychological Review, 90(316-338).

Murphy, G. L. (1984). Establishing and accessing referents in discourse. Memory & Cognition, 12(5), 489-497.

Nicol, J. L., & Pickering, M. J. (1993). Processing of syntactically ambiguous sentences: Evidence from semantic priming. Journal of Psycholinguistic Research, 22, 207-237.

Nicol, J. L., & Swinney, D. A. (1989). The role of structure in coreference assignment during sentence comprehension. Journal of Psycholinguistic Research, 18, 5-19.

Nobel, P. A., & Shiffrin, R. M. (2001). Retrieval processes in recognition and recall. Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(2), 384-413.

Nordlie, J., Dopkins, S., & Johnson, M. (2001). Words in a sentence become less accessible when an anaphor is resolved. Memory & Cognition, 29(2), 355-362.

O'Brien, E. J., Albrecht, J. E., Hakala, C. M., & Rizella, M. L. (1995). Activation and suppression of antecedents during reinstatement. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(3), 626-634.

O'Brien, E. J., Duffy, S. A., & Myers, J. L. (1986). Anaphoric inference during reading. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12(3), 346-352.

Pike, R. (1984). A comparison of convolution and matrix distributed memory systems. Psychological Review, 91, 281-294.

Poizner, H., Bellugi, U., & Tweney, R. D. (1981). Processing of formational, semantic, and iconic information in American Sign Language. Journal of Experimental Psychology: Human Perception and Performance, 7, 1146-1159.

Pollard, C. J., & Sag, I. A. (1994). Head-driven phrase structure grammar. Chicago: University of Chicago Press.

Pritchett, B. L. (1991). Head position and parsing ambiguity. Journal of Psycholinguistic Research, 20(3), 251-270.

Pritchett, B. L. (1992). Grammatical Competence and Parsing Performance. Chicago: University of Chicago.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981a). Order effects in recall. In A. Long & A. Baddeley (Eds.), Attention and performance (vol. 9) . Hillsdale, NJ: Erlbaum.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981b). Search of associative memory. Psychological Review, 88, 93-134.

Ratcliff, R., & Murdock, B. B., Jr. (1976). Retrieval processes in recognition memory. Psychological Review, 83, 190-214.

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

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.

Rochon, E., Waters, G. S., & Caplan, D. (1994). Sentence comprehension in patients with Alzheimer's disease. Brain & Language, 46(329-349).

Sanford, A. J., & Garrod, S. (1989). What, when and how? Questions of immediacy in anaphoric reference resolution. Language and Cognitive Processes, 4, 235-262.

Sanford, A. J., Moar, K., & Garrod, S. (1988). Proper names as controllers of discourse focus. Language and Speech, 31, 43-56.

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

Shieber, S. M. (1986). An introduction to unification-based approaches to grammar. Stanford, CA: Center for the Study of Language and Information, Stanford University.

Shiffrin, R. M., Ratcliff, R., & Clark, S. (1990). List-strength effect: II. Theoretical mechanisms. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 179-185.

Simon, H. A., & Zhang, G. (1985). STM capacity of Chinese words and idioms: Chunking and the acoustical loop hypothesis. Memory and Cognition, 13, 193-201.

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

Stine-Morrow, E. A. L., Ryan, S., & Leonard, J. S. (2000). Age differences in on-line syntactic processing. Experimental Aging Research, 26, 312-322.

Stromswold, K., Caplan, D., Alpert, N., & Rauch, S. (1996). Localization of syntactic comprehension by Positron Emission Tomography. Brain and Language, 52, 452-473.

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.

Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. Journal of Experimental Psychology, 86, 255-262.

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

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.

Trueswell, J. C., & Tannenhaus, M. K. (1994). Toward a lexicalist framework of constraint-based syntactic ambiguity resolution. In J. C. Clifton & L. Frazier (Eds.), Perspectives on sentence processing . Hillsdale, NJ: Erlbaum.

Tulving, E. (1983). Elements of episodic memory. New York: Oxford University Press.

Tulving, E., & Thomson, D. M. (1971). Retrieval processes in recognition memory: Effects of associative context. Journal of Experimental Psychology, 87, 175-184.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. Psychological Review, 80(5), 352-373.

Van Dyke, J. A., & Lewis, R. L. (submitted). Distinguishing effects of structure and decay on attachment and repair: A retrieval interference theory of recovery from misanalyzed ambiguities. Journal of Memory and Language.

Vos, S. H., Gunter, T. C., Schriefers, H., & Friederici, A. (2001). Syntactic parsing and working memory: The effects of syntactic complexity, reading span, and concurrent load. Language and Cognitive Processes, 16(1), 65-103.

Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. Cognition, 75, 105-143.

Walker, H. A., & John, E. E. (1984). Interference and facilitation in short-term memory for odors. Perception & Psychophysics, 36, 508-514.

Wanner, E., & Maratsos, M. (1978). An ATN approach in comprehension. In M. Halle, J. Bresnan, & G. Miller (Eds.), Linguistic Theory and Psychological Reality (pp. 119-161). Cambridge, MA: MIT Press.

Warren, T., & Gibson, E. (2000). The effects of discourse status on linguistic complexity. Paper presented at the 13th annual CUNY conference on human sentence processing., San Diego, CA.

Warren, T., & Gibson, E. (in press). The influence of referential processing on sentence complexity. Cognition.

Waters, G. S., & Caplan, D. (1996a). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). Psychological Review, 103(4), 761-772.

Waters, G. S., & Caplan, D. (1996b). Processing resource capacity and the comprehension of garden path sentences. Memory and Cognition, 24(3), 342-355.

Waters, G. S., Caplan, D., & Hildebrandt, N. (1991). On the structure of verbal short-term memory and its functional role in sentence comprehension: Evidence from neuropsychology. Cognitive Neuropsychology, 8(2), 81-126.

Waters, G. S., Caplan, D., & Rochon, E. (1995). Processing capacity and sentence comprehension in patients with Alzheimer's disease. Cognitive Neuropsychology, 12, 1-30.

Wiley, J., Mason, R. A., & Myers, J. L. (2001). Accessibility of potential referents following categorical anaphors. Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(5), 1238-1249.

Williams, H. L., Beaver, W. S., Spence, M. T., & Rundell, O. H. (1969). Digital and kinesthetic memory with interpolated information processing. Journal of Experimental Psychology, 80, 530-536.

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

Zurif, E. B., D., S., Prather, P., Solomon, J., & Bushell, C. (1993). An on-line analysis of syntactic processing in Broca's and Wernicke's aphasia. Brain & Language, 45, 448-464.

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. Journal of Experimental Psychology: Learning, Memory, & Cognition, 21, 386-397.