Journal of Experimental Psychology:
Learning, Memory, and Cognition
2007, Vol. 33, No. 2, 407–430

0278-7393/07/$12.00 DOI: 10.1037/0278-7393.33.2.407

Interference Effects From Grammatically Unavailable Constituents During Sentence Processing

Julie A. Van Dyke
Haskins Laboratories

Evidence from 3 experiments reveals interference effects from structural relationships that are inconsistent with any grammatical parse of the perceived input. Processing disruption was observed when items occurring between a head and a dependent overlapped with either (or both) syntactic or semantic features of the dependent. Effects of syntactic interference occur in the earliest online measures in the region where the retrieval of a long-distance dependent occurs. Semantic interference effects occur in later online measures at the end of the sentence. Both effects endure in offline comprehension measures, suggesting that interfering items participate in incorrect interpretations that resist reanalysis. The data are discussed in terms of a cue-based retrieval account of parsing, which reconciles the fact that the parser must violate the grammar in order for these interference effects to occur. Broader implications of this research indicate a need for a precise specification of the interface between the parsing mechanism and the memory system that supports language comprehension.

Keywords: cue-based memory retrieval, interference, complexity effects, sentence processing, constraint-based parsing

Interference effects in sentence processing are beginning to be recognized, but the conditions that give rise to these effects are not well understood (Gordon, Hendrick, & Johnson, 2001, 2004; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006). Gordon and colleagues (2001, 2004) observed that the classic processing advantage for subject-relative clauses over object-relative clauses was reduced or eliminated when the second noun phrase (NP) in sentences such as 1a and 1b was either a pronoun (*you* or *everyone*) or a proper name (*Joe*). They attributed their result to a reduction in similarity-based interference that occurs when the two NPs have different referential characteristics (i.e., common NPs refer via their description; pronouns and proper names refer directly to objects previously established in the discourse). Thus, on this account, interference effects are due to the presence of NPs with shared referential characteristics.

(1a) The banker that praised [the barber/a barber/Joe/you/everyone] climbed the mountain.

(1b) The banker that [the barber/a barber/Joe/you/everyone] praised climbed the mountain.

Experiment 1 was part of the author's doctoral dissertation, submitted to the Psychology Department at the University of Pittsburgh. Portions of this work were conducted while the author was supported on a University of Michigan dissertation fellowship, National Institutes of Health (NIH) Training Grant T32-HD-07548 to Haskins Laboratories, and NIH National Research Service Award F32-HD049215-0.

I thank Julie Fiez, Lyn Frazier, Ted Gibson, Rick Lewis, Brian McElree, Alice Healy, Chuck Perfetti, Eric Reichle, Whit Tabor, and Shravan Vasishth for helpful discussions of earlier versions of this work. Any errors or faults of clarity remain my own.

Correspondence concerning this article should be addressed to Julie A. Van Dyke, Haskins Laboratories, 300 George Street, New Haven, CT 06511-6695. E-mail: jvandyke@haskins.yale.edu

Several researchers have suggested an alternative account, one that implicates *retrieval* as the source of interference effects (Lewis, Vasishth, & Van Dyke, 2006; McElree, Foraker, & Dyer, 2003; Van Dyke, 2002; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006). Although these are not the first sentence-processing theories to include a retrieval component, previous theories (e.g., Gibson, 1998, 2000) have emphasized *decay* as the source of processing complexity, on the basis of hypothesized memory demands that certain structures present for comprehenders. In contrast, approaches that focus on *interference* as the primary determinant of complexity have drawn on specific retrieval mechanisms whose properties have been well studied in the memory literature. According to these theories, grammatical relations are created via cue-based retrieval of necessary constituents. Grammatical heads provide retrieval cues that identify necessary properties of the required constituent (e.g., grammatical case, thematic role, semantic properties), which are then combined in parallel, to create a single retrieval probe. One formalization of this idea is given in Equation 1, where the probability, P, of retrieving a particular item, I_i, to serve as the dependent of the cuing constituent is defined as the strength of the association, S, between each feature of the probe ($Q_1, \ldots, Q_m$), and the features of the actual memory trace ($I_1, \ldots, I_n$), denoted as $S(Q_j,I_i)^{w_j}$, where w_j is a weighting factor denoting the relative saliency of the different cues.

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

This equation was borrowed from mathematical models of memory retrieval (especially Gillund & Shiffrin, 1984; see also Hintzman, 1984, 1988; Nairne, 1990; Raaijmakers & Shiffrin, 1981) and is consistent with the adaptive control of thought—rational (ACT–R) computational model (cf. Lewis & Vasishth, 2005, for an ACT–R model of cue-based parsing). It is based on a large body of experimental data showing that memory retrieval depends both on the match between the retrieval probe and the retrieval target, as well as the match between the retrieval probe and all other items in memory. Specifically, it states that the probability of retrieving a particular item with a given retrieval probe is an increasing function of the probe-to-item strength and a decreasing function of the sum of the probe-to-item strengths for all other items stored in memory. Thus, similarity-based interference arises in the presence of one or more distractor items that are very similar to the target item vis-à-vis the retrieval probe, because its probe-to-item strength will be highly similar to the probe-to-target strength.

This framework offers a general explanation of interference effects arising both from the presence of similar items in memory, as suggested by the Gordon et al. (2001, 2004) data, and from ambiguous retrieval cues. In the former case, a common noun distractor that overlaps with many of the properties of the target noun (including grammatical, semantic, or referential) will increase the value of the denominator in Equation 1 as compared with when there is less overlap of a distractor with the target, as may be the case with a pronoun or proper noun distractor, whose semantic and referential properties may distinguish it from the target. This will make the probability of retrieving the target greater when there is a pronoun or proper name distractor (which supplies less interference) than when there is a common noun distractor, despite the fact that the retrieval cues (and hence the numerator) remain constant in the two cases.

Direct evidence in support of the effect of ambiguous retrieval cues was presented by Van Dyke and McElree (2006), who manipulated the probe-to-item match between the target item and its distractors. Using a memory load paradigm, where participants had to memorize a list of three words prior to reading a sentence, they compared conditions as in Sentence 2a, where the only NP to fit the semantic requirements for the object of the verb *sailed* is *boat*, and conditions as in Sentence 2b, where any of the NPs in memory (i.e., *table*, *sink*, *truck*, *boat*) could serve as the object of *fixed*.

(2a) TABLE–SINK–TRUCK
It was the boat that the guy who lived by the sea sailed in two sunny days.

(2b) TABLE–SINK–TRUCK
It was the boat that the guy who lived by the sea fixed in two sunny days.

The authors observed increased reading times at the manipulated verb for the interfering conditions like in Sentence 2b, an effect that disappeared when participants read these sentences without having to first memorize the distractor items. These results clearly implicate the role of retrieval, arising from retrieval cues that do not uniquely identify the object of *fixed* when there are other *fixable* items in memory, causing a reduction of the numerator in Equation 1 despite identical distractors across the two conditions. Notably, the referential similarity account presented by Gordon and colleagues (2001, 2004) cannot explain these effects, because the referential properties of the distractor items are identical (i.e., all descriptive NPs).

Additional evidence for the centrality of retrieval was provided by Van Dyke and Lewis (2003), who compared sentences such as 3a and 3b, in which a distracting NP intervenes between a long-distance subject–verb dependency. In these sentences, the retrieval probe comes from the phrase *was complaining*, in which the long-distance subject *resident* must be retrieved despite a more recent NP (*warehouse*). Increased reading times were observed in Sentence 3b as compared with Sentence 3a, despite intervening regions of the same length. Van Dyke and Lewis characterized 3b as a sentence containing *syntactic interference*, after the type of retrieval cue matching the distractor. Thus, in 3b, where the intervening NP is a grammatical subject, interference is produced by the match with the retrieval cues of *was complaining*, which is looking for its subject. In contrast, the intervening NP in 3a is the object of a preposition and hence does not match the subject cues of the verb.

(3a) The worker was surprised that the resident who was living near the dangerous warehouse was complaining about the investigation.

(3b) The worker was surprised that the resident who said that the warehouse was dangerous was complaining about the investigation.

The current experiments extend the approach of investigating interference produced by specific types of retrieval cues. Specifically, the role of *semantic interference* was examined, to determine whether items that fit the semantic retrieval cues from the retrieval probe could create difficulty. The evidence from Van Dyke and McElree (2006) described above suggests that they can; however, these effects were observed in the absence of any grammatical marking on the distracting items. If syntactic properties play a primary role in determining a sentence's interpretation, as proposed by syntax-first parsing models (Ferreira & Clifton, 1986; Frazier, 1978, 1990; Frazier & Rayner, 1982), then we may expect that semantic interference would not occur when the syntactic properties of the distracting items do not fit the syntactic retrieval cues from the retrieval probe. Thus, in a variation of Sentence 3a, given in Sentence 4, where the intervening NP *neighbor* fits the semantic cues of the verb *was complaining*, the semantic match of the NP is not predicted to create difficulty because *neighbor* is grammatically unavailable—it has already been assigned case marking as the object of a preposition and hence does not match the syntactic cues from *was complaining*, which is looking for a subject.

(4) The worker was surprised that the resident who was living near the dangerous *neighbor* was complaining about the investigation.

Evidence that semantic interference *would* occur has been observed by Tabor, Galantucci, and Richardson (2004), who demonstrated effects of local coherence in processing reduced relative clauses over and above the difficulty associated with processing the reduction itself. In particular, they found that Sentence 5a produced slower reading times and reduced grammaticality judgments as compared with its unreduced control than Sentence 5b when compared with an analogous unreduced control (local coherence shown in italics).

(5a) The coach smiled at *the player tossed a Frisbee* by the opposing team.

(5b) The coach smiled at the player thrown a Frisbee by the opposing team.

The authors interpreted their results as support for a constraint–satisfaction approach to parsing in which locally consistent fragments can create competition for the global parse (MacDonald, Pearlmutter, & Seidenberg, 1994; Stevenson, 1994, 1998; Tabor & Hutchins, 2004; Vosse & Kempen, 2000). According to this view, syntactic and semantic factors contribute simultaneously to produce the parse most consistent with the evidence from the input. One contribution of the current study is to provide evidence that these effects are not contingent on adjacency or local coherence. The cue-based retrieval account predicts that semantic interference should also be observed in a sentence like 6, where the distracting NP *neighbor* is distant from the verb *was complaining* as well as having no locally coherent grammatical analysis. This is also a prediction of self-organizing constraint–satisfaction models, such as Tabor and Hutchins' (2004) self-organizing parser (SOPARSE), which has no mechanism for restricting attachments to only adjacent items.

(6) The worker was surprised that the resident who said that the *neighbor* was dangerous was complaining about the investigation.

Broad effects of semantic interference in sentence processing have been previously observed but have not been investigated systematically using online processing measures. For example, Stolz (1967) observed that participants had more difficulty paraphrasing the clauses of center embedded sentences in which all of the NPs were potential subjects for the embedded verbs, as in 7a, as compared with sentences in which the meaning of the verbs helped to distinguish the more appropriate subject, as in 7b.

(7a) The chef that the waiter that the busboy appreciated teased admired good musicians.

(7b) The bees that the hives that the farmer built housed stung the children.

Similarly, King and Just (1991) found that a manipulation of pragmatic bias of verbs vis-à-vis the NPs in object-relative clauses affected comprehension accuracy for both high- and low-memory-capacity participants. Specifically, they found that participants had more difficulty answering true–false probes about the main clause in sentences like 8a, in which both NPs can serve as subject for the main verb, as compared with 8b, in which the meaning of the verbs can be used to identify the subject of each. King and Just also collected reading times for these sentences, but they reported only a weak effect of pragmatic bias on reading times ($.05 < p < .10$).

(8a) The robber that the fireman detested watched the program.

(8b) The robber that the fireman rescued stole the jewelry.

The experiments reported here are intended to replicate the syntactic interference effects observed earlier and to test the hypothesis that semantic interference effects will occur whenever a semantically suitable NP occurs in the region intervening between the subject and verb of a long-distance dependency. To accomplish this, I conducted three experiments in which syntactic and semantic interference were crossed in a 2 × 2 design, creating the four conditions in Table 1. The four conditions can be constructed by combining the sentence introduction and conclusion with one of the four intervening regions. To increase readability, I hereafter refer to the low and high *syntactic* interference conditions (as in Sentences 3a and 3b), respectively, as *LoSyn* and *HiSyn*. The low and high *semantic* interference conditions are called *LoSem* and *HiSem*.

Table 1
Example Syntactic and Semantic Interference Stimuli for Experiment 1 With Regions for Analysis

Sentence region	Example stimulus
Introduction	The worker was surprised that the resident
Intervening region	
LoSyn/LoSem	who was living near the dangerous warehouse
LoSyn/HiSem	who was living near the dangerous neighbor
HiSyn/LoSem	who said that the warehouse was dangerous
HiSyn/HiSem	who said that the neighbor was dangerous
Critical region	was complaining
Spillover region	about the
Last word	investigation

Note. LoSyn and HiSyn refer to low and high syntactic interference conditions, respectively; LoSem and HiSem refer to low and high semantic interference conditions, respectively.

The key claim of the retrieval account is that processing difficulty will be encountered whenever the retrieval cues from the verb *was complaining* do not unambiguously identify its subject. Thus, the HiSyn conditions are expected to be more difficult than the LoSyn conditions because of the intervening subject, which creates syntactic interference by providing a distracting NP with subject marking to match the subject cues from the retrieval probe. Similarly, the HiSem conditions are expected to be more difficult than the LoSem conditions because the intervening NP creates semantic interference due to *neighbor* being the type of entity that "can complain," hence matching the semantic cues from the retrieval probe.

These sentences were investigated using two different experimental paradigms: a moving-window paradigm with a "Got it?" task and an eye-tracking paradigm with cloze comprehension questions, which required participants to report the interpretation they had constructed during reading. It is expected that these paradigms will provide converging evidence for the observed effects, as well as yielding increasingly detailed information about how the syntactic and semantic interference effects operate during sentence processing.

Experiment 1: "Got It?" Task

A 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 participants to make explicit decisions about grammaticality, a topic that can cause considerable anxiety in otherwise capable students. In this task, participants are instructed to judge as quickly as possible after the end of a sentence whether they understood it. If the person had no difficulty understanding the sentence, he or she will answer "yes." A "no" answer indicates that the person was unable to make sense of the sentence, perhaps owing to difficulty completing the long-distance attachment. It is important to note that participants are encouraged not to attempt to make sense of an awkward sentence, which means that this task allows an estimate of the immediate effects of retrieval interference, prior to any deliberate attempts to make sense of the sentence.

Method

Participants

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

Materials

Forty-eight sets of experimental items were randomly chosen from the piloted set (described below) for use in Experiment 1. Four lists of items were constructed so that each participant received one of the four conditions from each set of items but no participant saw more than one condition in each set. Each participant received sentences in each of the four conditions, permitting a within-subject analysis of the data. Items were presented in a blocked random order such that every experimental item was separated from the next experimental item by three filler items of different syntactic constructions.

The filler items were designed to be appropriate matches for several aspects of the experimental sentences. 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 items were included to discourage participants from focusing on length as indicating a hard-to-comprehend sentence.

After this set of 144 filler items was constructed, half of the items were made ungrammatical via the addition of one or two words at random points in the sentences (e.g., *The friendly manager encouraged the employees earn with sizeable bonuses*). These items were included in order to maintain participants' vigilance in the "Got it?" task.

Piloting

The semantic interference conditions required piloting to ensure that the intervening noun could serve as a plausible subject for the final verb phrase. This involved transforming the semantic interference sentences into the three conditions presented in Table 2.

Table 2
Sentences for Semantic Interference Pilot

Condition	Sentence
Target	The worker was surprised that the resident was complaining about the investigation.
Implausible distractor	The worker was surprised that the warehouse was complaining about the investigation.
Plausible distractor	The worker was surprised that the neighbor was complaining about the investigation.

These sentences provide an appropriate test of semantic interference because semantic interference occurs only 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 *target* conditions in Table 2 are rated similarly, then we would expect to observe semantic interference when these NP–verb combinations appear in the structures in Table 1. Likewise, for semantic interference to be absent, the *implausible* condition must be recognized as such and be significantly different from both the plausible and the target condition.

So there would be a sufficient number of high-contrasting experimental items, a pool of 160 item sets were constructed following the paradigm illustrated in Table 1. The three piloting conditions discussed above (shown in Table 2) produced 480 sentences. The full set of 480 sentences was randomized and then divided in half so that any one participant would be required to rate only 240 sentences in a 1-hr testing period. In some cases, two out of the three sentences in a set occurred in the same list, but these did not occur sequentially. The sentences were presented on personal computers using the MEL Professional experimental package (Version 2; Schneider, 1995). Students from the University of Pittsburgh undergraduate psychology subject pool participated in the piloting experiment in exchange for partial course credit. Sixteen participants rated the first half of the stimuli, and 15 different participants rated the second.

Participants were instructed to rate each of the sentences on a 5-point "sensibility" scale (1 = *makes no sense*, 5 = *makes perfect sense*). During the experiment, each sentence appeared on the screen in its entirety; participants entered their rating for that sentence and then pressed the space bar to move to the next sentence. Participants had no time pressure and received no feedback in the experiment, and no filler items were included.

Two one-way analyses of variance (ANOVAs) were conducted on each item separately. The first compared the target and plausible distractor conditions, and no significant difference was expected between these two. Fifty-two of the 160 items did produce a significant difference between the two plausible conditions and were either dropped from the set or corrected, if an obvious alternative NP suggested itself. The second ANOVA compared the plausible and the implausible conditions, and we did expect to find a significant difference between these conditions. Five of the 160 item sets did not meet this criterion and were discarded or corrected. All corrected items were piloted with 12 participants in a more informal paper-test format, and all reached the same significance criteria described above except for 10 sets, which were dropped from the pool. The materials for Experiments 1–3 were randomly drawn from the final 150 item sets.

Procedure

As with the pilot experiment, Experiment 1 was implemented in the MEL Professional experimental package (Version 2; Schneider, 1995) and was run on personal computers. The 192 sentences

(48 experimental; 144 filler) were presented in a noncumulative, self-paced, moving-window format, where each sentence was presented one word at a time. Prior to the experiment, participants were instructed to answer "yes" or "no" to the question "Did you get it?" Following Frazier et al. (1983), we encouraged them to answer as quickly as possible, without trying to make sense of a sentence that sounded awkward. Participants were warned that some sentences in this experiment were designed to be difficult to understand. A series of six practice sentences were presented prior to the experiment so that participants could familiarize themselves with the keyboard and the presentation sequence.

Design and Analysis

This experiment produced two dependent measures: accuracy for the "Got it?" task and reading times from the self-paced presentation. Both measures were analyzed via a 2 (high or low syntactic interference) × 2 (high or low semantic interference) factorial repeated measures ANOVA using error terms based on participant (F_1) and item (F_2) variability. The results of these ANOVAs are presented together with the min-F' statistic (Clark, 1973). For comparisons between condition means, 95% confidence intervals (CIs) are reported, calculated using the mean squared error (MSE) of the associated effects from the participant analyses according to the procedure for within-participant CIs described by Loftus and Masson (1994; Masson & Loftus, 2003). All reported means are based on analyses with participants as the random factor. Only data from trials in which participants answered "yes" to the "Got it?" question were included in the reading time analysis. These methods and conventions for presentation are followed throughout the article.

For the reading time measure, three regions of interest were identified (see Table 1). The first is the critical region, containing the verb for the long-distance dependency. The second is referred to as the spillover region, containing the two words following the critical region. The last region contains only the last word of the sentence. Although it is common practice not to analyze the last word in sentence-processing experiments because of the expectation that reading times in this position are confounded with "sentence wrap-up" effects (Just & Carpenter, 1980; King & Just, 1991), we hypothesized that at least some of these sentence wrap-up effects may be related to the interference manipulations. In particular, if participants are distracted by the interfering NP, then incorrect attachments may not be identified until final interpretive processing is done at the end of the sentence.

All regions were identical for all conditions, and consequently no length corrections (i.e., number of characters) were performed on the data. Reading times were trimmed to within 2.5 times the standard deviation for each condition, and extreme times were replaced with the cutoff value. This affected 2.5% of the data.

Results

Table 3 presents both the proportion of the sentences in each condition for which participants said that they did "Get it" and reading time results. Table 4 presents the results of 2 × 2 within-subject ANOVA testing on the four interfering conditions. For brevity, the table reports results only for regions with at least one significant F statistic.

Table 3
Mean Accuracy and Reading Times in "Got It?" Task in Experiment 1, With Participants as the Random Factor

		Reading time (ms)		
Interference type	Accuracy	Critical region	Spillover region	Last word
LoSyn/LoSem	.91 (.02)	858 (24)	566 (16)	723 (43)
LoSyn/HiSem	.83 (.03)	912 (34)	571 (16)	697 (41)
HiSyn/LoSem	.81 (.03)	871 (30)	551 (18)	667 (37)
HiSyn/HiSem	.78 (.03)	875 (30)	568 (15)	822 (81)

Note. Values in parentheses are standard errors. LoSyn and HiSyn refer to low and high syntactic interference conditions, respectively; LoSem and HiSem refer to low and high semantic interference conditions, respectively.

For the "Got it?" measure, a main effect of syntactic interference was observed, with the HiSyn sentences being more difficult than the LoSyn sentences (.79 vs. .87; CI = .04). Similarly, a main effect of semantic interference was found, with the HiSem sentences being more difficult than the LoSem sentences (.80 vs. .86, CI = .03). The interaction was not significant. Nevertheless, because the local coherence hypothesis predicts an effect of semantic interference only in the LoSyn conditions, simple effects of semantic interference were tested. The data show that the effect of semantic interference was 8% in the LoSyn conditions and 3% in the HiSyn conditions. The effect was significant in the LoSyn conditions, $F_1(1, 34) = 13.65$, $p < .001$, $MSE_1 = 0.016$; $F_2(1, 47) = 5.61$, $p < .02$, $MSE_2 = 0.051$, but not in the HiSyn conditions (Fs < 1).

For the analysis of reading times in the critical region, the main effect of syntactic interference was not significant ($F < 1$). A significant main effect of semantic interference was observed, with the HiSem conditions being slower than the LoSem conditions (893 ms vs. 864 ms; CI = 23). The interaction was not significant. However, as with the "Got it?" judgments and as predicted by the local coherence account, the semantic interference effect was much larger in the LoSyn conditions (54 ms), $F_1(1, 34) = 6.85$, $p < .02$, $MSE_1 = 14{,}601$; $F_2(1, 46) = 4.24$, $p < .05$, $MSE_2 = 33{,}049$, than in the HiSyn conditions (4 ms; Fs < 1, *ns*).

No significant effects were observed in the spillover region (all Fs < 1). Reading times for the last word revealed no effect of syntactic interference (Fs < 1). The effect of semantic interference was marginal in the analysis by participants but significant in the analysis by items. This is due to the HiSem conditions being read more slowly than the LoSem conditions (789 ms vs. 701 ms; CI_2 = 77). The interaction was significant, $F_1(1, 34) = 5.39$, $p < .03$; $F_2(1, 46) = 7.14$, $p < .02$. Unlike the case for "Got it?" judgments and for reading times in the critical region, this interaction reflected a greater effect of semantic interference in the HiSyn conditions (155 ms), $F_1(1, 34) = 4.93$, $p < .04$, $MSE_1 = 170{,}674$; $F_2(1, 46) = 8.34$, $p < .01$, $MSE_2 = 211{,}618$, than in the LoSyn conditions (–27 ms; Fs < 1).

Discussion

The effect of syntactic interference observed in Van Dyke and Lewis (2003) was replicated in the offline "Got it?" task data, with HiSyn conditions receiving a lower proportion of positive re-

Table 4
Experiment 1: Analysis of Variance Results for All Dependent Measures

Measure	Main effect: Syntactic interference	Main effect: Semantic interference	Interaction
"Got it" response	$F_1(1, 34) = 14.14, p < .001, MSE_1 = 0.015$	$F_1(1, 34) = 9.96, p < .003, MSE_1 = 0.010$	$F_1 = 2.41$, *ns*
	$F_2(1, 47) = 8.87, p < .005, MSE_2 = 0.036$	$F_2(1, 47) = 5.61, p < .02, MSE_2 = 0.022$	$F_2 = 1.0$, *ns*
	min$F'(1, 81) = 5.45, p < .02$	min$F'(1, 80) = 3.59, p < .07$	
Reading time			
Critical region	$F_1 < 1$, *ns*	$F_1(1, 34) = 6.33, p < .02, MSE_1 = 4{,}636$	$F_1 = 2.36$, *ns*
	$F_2 < 1$, *ns*	$F_2(1, 46) = 4.11, p < .05, MSE_2 = 13{,}336$	$F_2 = 1.57$, *ns*
		min$F'(1, 81) = 2.49, p < .12$	
Last word	$F_1 < 1$, *ns*	$F_1(1, 34) = 3.06, p < .09, MSE_1 = 47{,}814$	$F_1(1, 34) = 5.39, p < .03, MSE_1 = 53{,}049$
	$F_2 > 1$, *ns*	$F_2(1, 46) = 5.22, p < .03, MSE_2 = 68{,}254$	$F_2(1, 46) = 7.14, p < .02, MSE_2 = 74{,}989$
		min$F'(1, 68) = 1.92, p < .17$	min$F'(1, 74) = 3.07, p < .09$

sponses than LoSyn conditions. This finding is consistent with the view that syntactic interference creates difficulty for making the correct attachment owing to the ambiguity of the available retrieval cues. However, the reading data in the critical region did not reveal an online effect. This finding is inconsistent with previous results and may be due to the loosely defined task; if participants were actually distracted by the intervening subject NP in the HiSyn conditions, there may be no reason for them to disturb their reading, because they erroneously believe that they have "gotten" the sentence.

The effect of semantic interference was observed in the offline "Got it?" data; however, in pairwise comparisons the effect was present only in the LoSyn conditions. This is consistent with the results of Tabor, Galantucci, and Richardson (2004), who observed that locally coherent dependencies that were inconsistent with a sentence's global interpretation create processing difficulty. They observed the effect both in reading times and in grammaticality judgments. The current data converge with theirs in that the semantic interference effect in the LoSyn constructions, in which the semantic manipulation created a local coherence with the critical verb, was also observed in slowed reading times at the critical region. The current data also extend their findings, in that the semantic interference effect was observed in the HiSyn conditions, where there was no local coherency. This effect was observed at the last word, which was later than the effect in the LoSyn conditions, suggesting that it may take longer for readers to realize they have created an incorrect interpretation when both the syntactic and the semantic properties of the distractor fit the retrieval cues from the verb.

Experiment 2: Reading Comprehension

The "Got it?" task from Experiment 1 provided preliminary evidence for both syntactic and semantic retrieval interference; however, the "Got it?" task makes the interpretation of the data somewhat problematic because there is no direct test of how well participants understood the sentences. To address this problem, the current experiment used comprehension questions as the offline measure to encourage participants to integrate incoming material into a consistent interpretation. The offline comprehension data are particularly important for evaluating the role of interference, as the essence of the effect is that an intervening NP will be incorrectly retrieved and interpreted as the subject of the cuing verb because of its match with the verb's retrieval cues. If evidence for this incorrect interpretation in offline comprehension results were found, this would not only support the retrieval interference account but also suggest that when incorrect retrievals occur they are not easily corrected. If, however, the effect is present only in online data, then it would appear that when incorrect retrievals occur, they can be reanalyzed online, resulting in only momentarily longer reading times, without lasting effects on the sentence's final interpretation.

A second aim of the current experiment is to seek online evidence for the syntactic interference effect. As noted above, Experiment 1 did not reveal this effect, despite the fact that the two LoSem conditions (LoSyn/LoSem and HiSyn/LoSem) were nearly identical to the conditions in Van Dyke and Lewis (2003), where the effect was first demonstrated. One explanation for this failure to replicate is that the task in Experiment 1 was quite different from that in the original study, which used self-paced reading with yes–no comprehension questions. In Experiment 2, we use an eye-tracking method that provides a highly naturalistic reading situation. If the previously observed effect was not an artifact due to a particular method, then we would expect it to be observed in the current experiment, especially in the LoSem conditions, which are free from any ambiguity associated with the semantics of the distracting NP.

Finally, the online data in Experiment 1 suggested that the semantic interference effect arises later in the HiSyn conditions (in the last region) than in the LoSyn conditions, where it occurs in the critical region. This is unexpected according to the retrieval account, because effects are expected at the point where the critical retrieval is made (i.e., the critical region). One possibility is that the later semantic effect arises because the strong (syntactic and semantic) fit of the incorrect NP to the verb causes participants to be "garden pathed" into believing they have created a coherent parse. In this case, they may detect the inconsistency only as a part of sentence-final wrap-up, where they discover that they have incorrectly interpreted a single NP in two incompatible grammatical roles. This may initiate attempts to reanalyze, resulting in slower reading times. Further data are necessary to clarify why it would be easier to initiate such reanalyses in the LoSyn conditions, hence producing an earlier effect. With Experiment 2, we seek to discover whether these timing differences were an artifact of the moving-window paradigm or whether they can be observed in the more natural reading context provided by use of an eye tracker.

The eye-tracking paradigm allows measures of both "early" processing and "later" processing, as well as a qualitative measure of reading difficulty (i.e., proportion of regressions back in order to reread portions of the sentence). Earlier processing is captured in first-pass reading times, which include all fixations within a region, starting with the first fixation until the reader's gaze exits the region. Later processing is characterized by regression path time and total time. Regression path includes all fixations from the first fixation in the region and all fixations in the current or in prior regions until the reader's gaze moves rightward out of the region (e.g., Brysbaert & Mitchell, 1996; Konieczny, Hemforth, Scheepers, & Strube, 1997). This is generally interpreted as the time needed to integrate a string before the reader is ready to process new material and may therefore include processing time associated with reinterpretation of incorrect dependencies. In cases where there are no leftward regressions, regression path time is equivalent to first-pass time. Total time includes both the first-pass fixation and all subsequent fixations in a region after the eyes have exited that region and returned, including rereading time originating from regions before or after the current region.

Method

Participants

Thirty-six participants from the New Haven, Connecticut, area were recruited via advertisements in local newspapers and fliers distributed around the city. They participated in this study as part of a larger study investigating individual differences in sentence processing and were paid $12.50/hr for 2.5 hr of testing (1.5 hr of reading skills testing and 1 hr of eye-tracking time). All of the participants were between the ages of 16 and 24 and had reading comprehension scores at the 12th-grade level or above on the even items from the Peabody Individual Achievement Test (Markwardt, 1998).[1] All were native speakers of American English, and none were university students.

Materials

Thirty-six sets of experimental items were chosen from the piloted pool discussed above, without regard for whether they had appeared in Experiment 1. Four lists of items were constructed, and each participant received only one of the four conditions from each set. The items were presented in blocked random order so that every experimental item was separated by three filler items, which were sentences from a different experiment. A total of 144 sentences were presented during the experiment.

After every experimental sentence and after half of the filler sentences, a comprehension question followed. The question was presented in a cloze format that tested the critical dependency of the associated experimental item with a two-alternative forced-choice decision. For example, in the sentences illustrated in Table 1, the question was "__ was complaining about the investigation." The correct answer for this example is *resident*, and possible distractors are *worker* (matrix subject) or *neighbor* (intervening noun). So that we could evaluate how often participants chose an incorrect noun as the subject of the critical verb, the LoSem conditions had *worker* as the distractor, as it was the only other semantically appropriate noun in the sentence. In the HiSem conditions, two semantically appropriate nouns occurred in the sentence (*worker* and *neighbor*), but *neighbor* was always presented as the distractor to allow us to evaluate whether participants would in fact interpret the intervening noun as the subject of the critical verb. These choices enabled a preliminary test of whether incorrect retrievals have specific consequences for comprehension. If participants were no more likely to choose this noun than the matrix noun, this might suggest a more general comprehension failure, one perhaps related to the complexity of the later portions of the sentence. Admittedly, this introduced a confound of the semantic manipulation and the foil for the comprehension question, an issue that was addressed in Experiment 3.

Procedure

Participants were seated in front of a 17-in. display with their eyes approximately 64 cm from the display. They wore an Eyelink II head-mounted eye tracker (SR Research, Mississauga, Ontario, Canada), sampling at a rate of 250 Hz from both eyes. Sentences were presented one at a time on a single line, with a maximum of 90 characters, using a monospace font. Type size was such that each character subtended about 17 min of visual arc. The eye tracker was calibrated using a series of nine fixed targets distributed around the display, followed by a 9-point accuracy test. Calibration was monitored throughout the experiment and was repeated after any breaks or whenever the experimenter judged necessary. Data were collected from both eyes, but analyses were done only on the right eye for all participants except one, whose right eye would not calibrate. Data from this participant's left eye were used for the analyses.

Prior to the experiment, participants were instructed to read each sentence for comprehension and told that they would be required to answer a comprehension question. Participants were also told that they could take a break at any time during the experiment. Each trial began with a screen containing a fixation point in the middle left of the display. While fixating on this point, participants were to press a button to bring up a sentence (the sentence would not appear unless participants fixated on the fixation point). After they had read the sentence, participants pressed the same button to view the comprehension question. The question appeared in the center of the screen; the two possible answers (the correct answer

[1] Odd-numbered items from the Peabody Individual Achievement Test were used to assess listening comprehension. All included participants were above the 12th-grade level on this test as well except one, who scored at the11.6th-grade level.

and the distractor) appeared three lines below, one to the left of center and one to the right of center. Participants indicated their answer by pressing the associated button on a button box; for example, if the answer appeared to the left of center they were to press the left button. The position of the correct answer was counterbalanced throughout the experiment. Participants were limited to 10 s for reading the stimulus sentence and 30 s for answering the comprehension question. If participants had not signaled that they had completed reading the sentence before the 10-s limit, the computer moved on to the comprehension question automatically. This occurred in less than 5% of trials. Participants were told to make their best guess at the comprehension question if they were unsure of the answer. If they had not answered within the 30-s limit, the computer moved on to the next item. This occurred in less than 1% of trials.

Data Analysis

All dependent measures were analyzed using a 2 (syntactic interference) × 2 (semantic interference) ANOVA. In addition to accuracy on the comprehension question, four eye-tracking measures (first pass, regression path, total reading time, and proportion of regressions back) are reported on the same three regions of interest analyzed in Experiment 1 (see Table 1). Data from accurate trials only were included in reading time analyses. As in previous experiments, analyses were done on raw reading times as the material was identical in all conditions. Fixations of less than 50 ms were not recorded, and any reading times greater than 2.5 times the standard deviation for that condition were replaced by the cutoff value. Across all reading time measures, this affected 2.4% of the data.

Results

Comprehension Questions

Results are presented in Table 5, and the results of ANOVA testing are presented in Table 6. The main effect of syntactic interference was significant both by participants and by items, with participants being more accurate on the LoSyn constructions than on the HiSyn constructions (.86 vs. .79; CI = .04). The main effect of semantic interference was also significant in both analyses, with the LoSem conditions being easier than the HiSem conditions (.88 vs. .77; CI = .05). The interaction was not significant. Nevertheless, pairwise comparisons of the effect of semantic interference in the LoSyn and HiSyn conditions were conducted separately, because the local coherence account predicts the effect in the LoSyn conditions only. This prediction was not supported in this experiment, as the effect of semantic interference was highly significant in both the LoSyn and HiSyn conditions. For LoSyn, the difference was 8%, $F_1(1, 35) = 9.45$, $p < .005$, $MSE_1 = 0.025$; $F_2(1, 35) = 11.34$, $p < .003$, $MSE_2 = 0.021$; for HiSyn, the difference was 13%, $F_1(1, 35) = 11.44$, $p < .003$, $MSE_1 = 0.05$; $F_2(1, 35) = 15.77$, $p < .001$, $MSE_2 = 0.039$.

Table 5
Mean Accuracy Scores for Comprehension Questions in Experiment 2, With Participants as the Random Factor

Interference type	Accuracy
LoSyn/LoSem	.90 (.02)
LoSyn/HiSem	.82 (.03)
HiSyn/LoSem	.86 (.03)
HiSyn/HiSem	.73 (.04)

Note. Values in parentheses are standard errors. LoSyn and HiSyn refer to low and high syntactic interference conditions, respectively; LoSem and HiSem refer to low and high semantic interference conditions, respectively.

Reading Time Measures

Table 7 summarizes the results for each measure in the regions of interest.

Critical region. Table 8 presents the results of ANOVA testing for all dependent measures in this region. The syntactic interference effect was significant for all four measures, and no other effects were significant. For the first-pass reading times, the mean reading time was 379 ms for LoSyn versus 416 ms for HiSyn (CI = 25). Tests for simple effects showed the effect to be significant for the LoSem conditions in the analysis with participants as the random factor, but not for items, $F_1(1, 35) = 4.82$, $p < .04$, $MSE_1 = 10{,}398$; $F_2 = 2.76$. The effect was marginal in the HiSem conditions in the analysis by participants, $F_1(1, 35) = 3.81$, $p < .06$, $MSE_1 = 12{,}387$; $F_2 = 2.08$.

For the regression path measure, the mean reading time was 475 ms for LoSyn versus 629 ms for HiSyn (CI = 59). In pairwise comparisons this effect held in both the LoSem conditions, $F_1(1, 35) = 15.23$, $p < .001$, $MSE_1 = 46{,}241$; $F_2(1, 35) = 6.49$, $p < .02$, $MSE_2 = 66{,}300$, and the HiSem conditions, $F_1(1, 35) = 15.38$, $p < .001$, $MSE_1 = 66{,}614$; $F_2(1, 35) = 15.50$, $p < .001$, $MSE_2 = 64{,}459$.

For total time in the region, the mean reading time was 642 ms for LoSyn and 750 ms for HiSyn (CI = 39). The pairwise comparisons showed the effect in both the LoSem conditions, $F_1(1, 35) = 10.32$, $p < .002$, $MSE_1 = 40{,}737$; $F_2(1, 35) = 6.43$, $p < .02$, $MSE_2 = 37{,}489$, and the HiSem conditions, $F_1(1, 35) = 15.34$, $p < .001$, $MSE_1 = 27{,}376$; $F_2(1, 35) = 8.84$, $p < .005$, $MSE_2 = 50{,}181$.

Participants made regressive eye movements backward from the critical region in 17% of trials (see Table 7). Despite this small number of regressions, the syntactic interference effect was observed (.13 for LoSyn vs. .20 for HiSyn; CI = .04). This was significant for the LoSem conditions, $F_1(1, 35) = 5.54$, $p < .03$, $MSE_1 = 0.022$, although not by items ($F_2 = 1.24$). The syntactic effect in the HiSem conditions was also significant, $F_1(1, 35) = 4.98$, $p < .04$, $MSE_1 = 0.053$; $F_2(1, 35) = 6.13$, $p < .02$, $MSE_2 = 0.054$.

Although the pattern of reading times was consistent with the semantic interference effect in all measures, the effect did not reach significance in any measure. There was a marginal effect ($p < .09$) in the regression path measure (524 ms for LoSem vs. 579 ms for HiSem; CI = 60). The interaction was not significant for any measure.

Spillover region. Table 9 presents the results of ANOVA testing in this region; only measures with significant effects are displayed. In the first-pass reading measure, the effect of syntactic interference was significant; however, the pattern of results was opposite to the expected effect (342 for LoSyn vs. 310 for HiSyn; CI = 24). This finding must be interpreted in the context of the

Table 6
Experiment 2: Analysis of Variance Results for Comprehension Questions

Measure	Main effect: Syntactic interference	Main effect: Semantic interference	Interaction
Accuracy	$F_1(1, 35) = 9.49, p < .005, MSE_1 = .016$ $F_2(1, 35) = 6.27, p < .02, MSE_2 = .025$ $\min F'(1, 67) = 3.78, p < .06$	$F_1(1, 35) = 20.35, p < .001, MSE_1 = .019$ $F_2(1, 35) = 26.68, p < .001, MSE_2 = .015$ $\min F'(1, 69) = 11.54, p < .002$	$F_1 < 1$, *ns* $F_2 = 1.51$, *ns*

significant crossover interaction in which the semantic interference effect was significant with means in the expected direction for the LoSyn conditions (320 ms for LoSyn/LoSem vs. 364 ms for LoSyn/HiSem; CI = 37), $F_1(1, 35) = 5.58, p < .02, MSE_1 = 12{,}625$; $F_2(1, 35) = 4.29, p < .05, MSE_2 = 22{,}762$, but not for the HiSyn conditions. Although the effect in the HiSyn conditions was not significant, $F_1(1, 35) = 3.05, p < .10, MSE_1 = 9{,}562$; $F_2(1, 35) = 1.57, MSE_2 = 8{,}571$, the pattern of means was in the direction opposite that predicted (325 ms for HiSyn/LoSem vs. 296 ms for HiSyn/HiSem; CI = 32). The presence of the effect in the LoSyn conditions is consistent with both the interference account and the local coherence account; however, neither account predicts the observed pattern of reading times in the HiSyn conditions. The absence of a significant effect in the HiSyn conditions is consistent with the local coherence account.

In the regression path measure, the effect of syntactic interference was significant (1,087 ms for LoSyn vs. 1,330 ms for HiSyn; CI = 137). Pairwise comparisons showed that the syntactic interference effect was statistically reliable in the LoSem conditions (395 ms), $F_1(1, 35) = 15.60, p < .001, MSE_1 = 360{,}080$; $F_2(1, 35) = 14.82, p < .001, MSE_2 = 472{,}362$, but not in the HiSem conditions (Fs < 1). The semantic interference effect was not significant. There was a trend for a crossover interaction in the analysis by participants ($p < .07$), but this effect did not near significance in the analysis by items. As predicted by the local coherence account, the effect of semantic interference was much greater in the LoSyn conditions (234 ms), $F_1(1, 35) = 9.31, p < .005, MSE_1 = 212{,}235$, although it did not reach significance in the analysis by items ($F_2 = 2.20$). The semantic interference effect in the HiSyn conditions was −70 ms, in the direction opposite that predicted, but this did not reach significance (Fs < 1). There were no significant effects for the total reading time measure. Overall, participants made regressions out of the spillover region on 57% of the trials; however, there were no significant effects of the experimental manipulations.

Final region. Table 10 presents the results of ANOVA testing in this region; as before, only measures with significant effects are displayed. There were no significant effects in the first-pass reading times at the last word. In the regression path measure there was a significant effect of syntactic interference (1,751 ms for LoSyn vs. 2,028 ms for HiSyn; CI = 222). Pairwise comparisons revealed that the effect was reliable in the HiSem conditions but only in the analysis by items, $F_1(1, 27) = 2.76$; $F_2(1, 35) = 3.98, p < .05, MSE_2 = 1{,}051{,}151$. The effect was not reliable in the LoSem conditions (Fs < 1.64). The effect of semantic interference was significant in the analysis by items (1,738 ms for LoSem vs. 2,022 ms for HiSem; CI = 235), but this effect did not reach significance in the analysis by participants (1,810 ms for LoSem vs. 1,969 ms for HiSem; CI = 214). The interaction was not significant. Although the local coherence account predicts no effect of semantic interference in the HiSyn conditions, the effect was numerically larger in the HiSyn conditions (206 ms) and statistically reliable in the analysis by participants (though not by items), $F_1 < 1.10$; $F_2(1, 35) = 4.92, p < .04, MSE_2 = 831{,}501$, and smaller in the LoSyn conditions (112 ms) and nonsignificant ($F_1 < 1$; $F_2 < 1.05$). There were no effects in the total time measure.

Participants regressed back from the final region in 87% of the trials. Table 7 shows that the syntactic interference manipulation produced more backward regressions, and this was significant in the analysis by items (.82 for LoSyn vs. .90 for HiSyn; CI = .07) but marginal in the analysis by participants (.84 for LoSyn vs. .89 for HiSyn; CI = .05, $p = .10$). The main effect of semantic interference was not significant. There was a trend for an interaction in the analysis by participants ($p < .08$), wherein fewer regressions were made when semantic interference was present in

Table 7
Experiment 2: Raw Reading Times (in Milliseconds) and Proportion of Regressive Eye Movements for Each Region for Each Dependent Measure, With Participants as the Random Factor

Measure and interference type	Critical region	Spillover region	Final word
First pass			
LoSyn/LoSem	376 (16)	320 (12)	286 (19)
LoSyn/HiSem	382 (19)	364 (21)	274 (19)
HiSyn/LoSem	413 (21)	325 (16)	259 (20)
HiSyn/HiSem	418 (19)	296 (15)	271 (22)
Regression path			
LoSyn/LoSem	454 (26)	970 (98)	1,695 (213)
LoSyn/HiSem	495 (30)	1,205 (100)	1,806 (183)
HiSyn/LoSem	594 (41)	1,365 (140)	1,925 (192)
HiSyn/HiSem	663 (44)	1,295 (140)	2,131 (244)
Total time			
LoSyn/LoSem	630 (35)	502 (30)	362 (36)
LoSyn/HiSem	653 (35)	540 (29)	373 (35)
HiSyn/LoSem	738 (42)	491 (25)	349 (34)
HiSyn/HiSem	761 (38)	493 (24)	360 (40)
Proportion of regressions			
LoSyn/LoSem	.12 (.02)	.54 (.05)	.81 (.05)
LoSyn/HiSem	.14 (.02)	.50 (.05)	.87 (.04)
HiSyn/LoSem	.18 (.02)	.60 (.05)	.92 (.03)
HiSyn/HiSem	.22 (.03)	.53 (.04)	.86 (.05)

Note. Values in parentheses are standard errors. LoSyn and HiSyn refer to low and high syntactic interference conditions, respectively; LoSem and HiSem refer to low and high semantic interference conditions, respectively.

Table 8
Experiment 2: Analysis of Variance Results for Reading Time Measures in the Critical Region

Measure	Main effect: Syntactic interference	Main effect: Semantic interference	Interaction
First pass	$F_1(1, 35) = 8.32, p < .01, MSE_1 = 5{,}847$ $F_2(1, 35) = 5.47, p < .03, MSE_2 = 5{,}391$ $\min F'(1, 67) = 3.30, p < .07$	$F_1 < 1.13$, *ns* $F_2 < 1$, *ns*	$F_1 < 1$, *ns* $F_2 < 1$, *ns*
Regression path	$F(1, 35) = 27.21, p < .001, MSE_1 = 31{,}490$ $F_2(1, 35) = 20.51, p < .001, MSE_2 = 33{,}402$ $\min F'(1, 69) = 11.69, p < .002$	$F_1(1, 35) = 3.24$, *ns*, $MSE_1 = 33{,}555$ $F_2(1, 35) = 3.12$, *ns*, $MSE_2 = 22{,}632$	$F_1 <1$, *ns* $F_2 < 1$, *ns*
Total time	$F_1(1, 35) = 29.95, p < .001, MSE_1 = 14{,}028$ $F_2(1, 35) = 12.24, p < .001, MSE_2 = 27{,}344$ $\min F'(1, 60) = 8.69, p < .005$	$F_1 < 1$, *ns* $F_2 < 1$, *ns*	$F_1 < 1$, *ns* $F_2 < 1$, *ns*
Regressive eye movements	$F_1(1, 35) = 10.71, p < .002,$ $MSE_1 = 0.017$ $F_2(1, 35) = 6.99, p < .01, MSE_2 = 0.024$ $\min F'(1, 67) = 4.23, p < .05$	$F_1 < 2.07$, *ns* $F_2 < 1.27$, *ns*	$F_1 < 1$, *ns* $F_2 < 1$, *ns*

the HiSyn conditions (.92 for LoSem vs. .86 for HiSem; CI = .05). In the LoSyn conditions, the semantic manipulation resulted in *more* regressions (.81 for LoSem vs. .87 for HiSem; CI = .09), but this difference was not significant. In addition, the syntactic manipulation had a greater effect in the LoSem conditions (.81 for LoSyn vs. .92 for HiSyn; CI = .09), $F_1(1, 27) = 4.72, p < .04, MSE_1 = 0.067$; $F_2(1, 30) = 4.84, p < .04, MSE_2 = 0.065$. The effect of the syntactic manipulation in the HiSem conditions (.87 for LoSyn and .86 for HiSyn; CI = .06) was not significant (*F*s < 1).

Discussion

This experiment is consistent with Experiment 1 in showing effects of both syntactic and semantic interference in the comprehension questions. Whereas the effect of semantic interference was confined to the LoSyn conditions in the previous experiment, there is evidence for the effect in both the LoSyn and the HiSyn conditions in the current experiment. This suggests that the results from Experiment 1 may have been influenced by the nature of the "Got it?" question, which did not directly query the interpretation participants had constructed. The cloze format used here asked participants to report which NP they had interpreted as the subject of the critical verb and so may have been a more authentic measure of participants' interpretation.

The effect of syntactic interference was observed in the reading times at the critical region in the current experiment. This is unlike the results in Experiment 1 but is consistent with results of previous experiments (Van Dyke & Lewis, 2003). The effect occurred in all measures in the critical region, beginning from the earliest measure (first-pass reading time), and was particularly strong in the LoSem conditions, which most closely replicate previous work. This finding suggests that the syntactic role alone is sufficient to create interference effects, as participants are apparently distracted by an intervening subject in the HiSyn conditions, even when its semantic properties make it unsuitable as a subject of the critical verb. Moreover, the effect occurs early—as soon as the critical retrieval occurs.

There was also evidence of syntactic interference in the regression path measure in the spillover region. Because there is no reason to suggest a delay in integrating the critical verb into the sentence, the slowdown in this region is likely due either to slowdown actually associated with the critical region or to early attempts to reanalyze an incorrect dependency formed by retrieving the incorrect noun in the critical region in the HiSyn/LoSem conditions. The latter possibility is consistent with the absence of an effect in the first-pass measure, which would be expected if this were an actual "spillover" effect, and the fact that the syntactic effect was present only in the LoSem conditions, which suggests that participants may not have realized they had been distracted by the interfering NP when it was both semantically and syntactically suitable as the subject of the critical verb. This explanation is also consistent with the reversed effect of syntactic interference in the first-pass measure in the spillover region, caused by faster reading

Table 9
Experiment 2: Analysis of Variance Results for Reading Time Measures in the Spillover Region

Measure	Main effect: Syntactic interference	Main effect: Semantic interference	Interaction
First pass	$F_1(1, 35) = 6.58, p < .02, MSE_1 = 5{,}471$ $F_2(1, 35) = 4.18, p < .05, MSE_2 = 9{,}888$ $\min F'(1, 67) = 2.57, p < .12$	$F_1 < 1$, *ns* $F_2 < 1$, *ns*	$F_1(1, 35) = 5.22, p < .03, MSE_1 = 9{,}107$ $F_2(1, 35) = 8.01, p < .01, MSE_2 = 5{,}728$ $\min F'(1, 67) = 3.16, p < .08$
Regression path	$F_1(1, 35) = 12.27, p < .001, MSE_1 = 172{,}795$ $F_2(1, 35) = 8.89, p < .005, MSE_2 = 434{,}864$ $\min F'(1, 69) = 5.16, p < .03$	$F_1 < 1$, *ns* $F_2 < 1$, *ns*	$F_1(1, 35) = 3.45$, *ns* $F_2(1, 35) = 2.22$, *ns*

Table 10
Experiment 2: Analysis of Variance Results for Reading Time Measures in the Final Region

Measure	Main effect: Syntactic interference	Main effect: Semantic interference	Interaction
Regression path	$F_1(1, 27) = 4.74, p < .04$, $MSE_1 = 453{,}349$	$F_1(1, 27) = 1.69$, *ns*, $MSE_1 = 419{,}003$	$F_1 < 1.00$, *ns*
	$F_2(1, 30) = 4.81, p < .04$, $MSE_2 = 537{,}625$	$F_2(1, 30) = 4.96, p < .04$, $MSE_2 = 506{,}614$	$F_2 < 2.22$, *ns*
	min$F'(1, 57) = 2.39$, *ns*	min$F'(1, 44) = 1.26$, *ns*	
Regressive eye movements	$F_1(1, 27) = 2.83$, *ns*, $MSE_1 = 0.023$	$F_1 < 1.00$, *ns*	$F_1(1, 27) = 3.47$, *ns* $MSE_2 = 0.027$
	$F_2(1, 30) = 4.49, p < .05$, $MSE_2 = 0.040$	$F_2 < 1.00$, *ns*	$F_2 < 1.00$, *ns*
	min$F'(1, 53) = 1.74$, *ns*		

of the HiSyn/HiSem conditions in this region. If both the syntactic and the semantic cues from the critical verb match properties of the incorrect noun, participants may believe they have successfully integrated the critical verb into the sentence, causing them to read through the following region more quickly.

The observation of semantic interference in the reading times is also consistent with this view. Unlike in Experiment 1, where the effect in the LoSyn conditions occurred in the critical region, here the effect occurred in the spillover region and was observed in both the early measure (first pass) and the regression path measure. The semantic interference effect in the HiSyn conditions occurred at the last word, as in Experiment 1, and was observed in Experiment 2 in both the regression path measure and the proportion of regressions, although not in the first-pass or total reading time measures. The overall pattern suggests that participants have an easier time noticing that they have been distracted by the intervening NP in the LoSyn conditions and attempt to correct their interpretation prior to the end of the sentence. In contrast, participants appear to be reading to the end of the sentence in the HiSyn conditions before making attempts to revise an incorrect interpretation. The extremely low accuracy rates for the HiSyn/HiSem conditions suggest that participants often fail to correct their interpretation, and this notion is consistent with the decrease in the proportion of regressions at the last word for this condition. Participants appear to be fooled by the distracting subject NP in the HiSyn conditions, particularly when it is semantically suitable as the subject of the critical verb, and may not realize that a correction is necessary. The regression path results indicate that when participants do notice the error, they spend substantially more time rereading than when the intervening NP is not a semantically suitable subject.

These conclusions about the results in the final region must be qualified by the loss of statistical power in the analysis by participants, which occurred because some participants did not fixate the last word of the sentence. This reduced the overall number of observations available for analysis and resulted in several effects reaching significance only in the analysis by items. Although it is common for participants to skip small regions during eye tracking, these regions were chosen in order to compare the location of effects in the current experiment with those observed in Experiment 1. One aim of Experiment 3 is to investigate whether the emerging pattern wherein the semantic interference effect in the LoSyn conditions occurs prior to the end of the sentence and the effect in the HiSyn conditions occurs primarily at the last word will be extended when the size of both the spillover region and the final region is larger. The retrieval account makes the prediction that difficulty associated with the retrieval itself will occur in the critical region. Effects arising later are likely associated with attempts to reanalyze incorrect dependencies created when an interfering NP was incorrectly retrieved in the critical region and interpreted as the subject of the critical verb. Additional explanation would be required to account for a differential time course for the semantic interference effect in LoSyn versus HiSyn constructions, should this prove to be a general pattern.

Experiment 3

One goal of the previous experiment was to replicate the online effect of syntactic interference observed in Van Dyke and Lewis (2003). Although this objective was achieved in Experiment 2, a possible alternative explanation is available for the slowdown observed in the reading times in the critical region. In all cases, the syntactic interference manipulation contained two adjacent verbs, with the verb of the embedded clause occurring just prior to the critical verb, creating the possibility that the elevated reading times at the critical verb were caused either by a "stumble" over that second verb or else by a spillover of slowed times from the first verb. The current experiment seeks to test this hypothesis with an eye-tracking experiment identical to that in Experiment 2 except that the two verbs are separated by an adverbial phrase positioned prior to the critical verb (see Table 11). If the syntactic interference effect is simply an artifact of reading two adjacent verbs, then it should not be present in the critical region in the current experiment. This also constitutes another test of whether local coherence is a necessary condition for the semantic interference effect, because the local coherence in the LoSyn conditions is now broken. If this were the cause for the difficulty caused by the semantic manipulation in the LoSyn conditions, then no difficulty should be observed here at all, even in offline comprehension measures. However, if the semantic interference effect arises whenever a suitable NP intervenes between the verbal retrieval probe and the target NP, then the effect should be observed clearly in both the LoSyn and the HiSyn conditions.

Table 11
Example Items for Experiment 3 With Regions for Analysis

Sentence region	Example item
Introduction	The pilot remembered that the lady
Intervening region	
LoSyn/LoSem	who was sitting in the smelly seat
LoSyn/HiSem	who was sitting near the smelly man
HiSyn/LoSem	who said that the seat was smelly
HiSyn/HiSem	who said that the man was smelly
Pre-critical region	yesterday afternoon
Critical region	moaned
Spillover region	about a refund
Final region	for the ticket

Note. LoSyn and HiSyn refer to low and high syntactic interference conditions, respectively; LoSem and HiSem refer to low and high semantic interference conditions, respectively.

Method

Participants

Forty undergraduates from New York University were recruited to participate in the study. They were paid $10/hr for 1 hr of testing. All were native speakers of American English.

Materials

The items from Experiment 2 were adapted for use in this experiment by inserting an adverbial phrase prior to the critical verb in all conditions. In some cases this required vocabulary changes to the items used in Experiment 2 to make the preposition fit more naturally into the sentence. To be certain that these changes did not affect the interfering properties of the distracting NP, we conducted a plausibility norming experiment on the new materials. For each experimental sentence, nine plausibility judgments were collected (see Table 12). Judgments 1–3 were of plausibility of each NP in the sentence as the subject of the critical cuing verb. Judgments 4 and 5 queried the plausibility of the target NP versus the distracting NP as subject of the critical verb. Judgments 6–9 evaluated how naturally the inserted preposition fit with the embedded clause from the experimental sentence. Table 12 contains the norming sentences derived from the experimental sentences in Table 11, together with the mean plausibility ratings for each sentence type (1 = *not plausible*; 7 = *highly plausible*).

Each of the 40 participants tested in the main experiment was invited to return to participate in the norming experiment in exchange for an additional $10. Of the 40 participants, 26 responded. Participants were asked to rate each of the nine derived sentences for each experimental item, together with 72 unrelated filler items. A within-subject ANOVA was conducted on Sentences 1–3 and revealed no significant difference (Fs < 1). Pairwise comparisons of the three sentences were all nonsignificant as well ($p > .27$). A within-subject ANOVA on Sentences 4 and 5 yielded the expected significant difference, $F_1(1, 25) = 332.30, p < .001, MSE_1 = 0.641; F_2(1, 35) = 453.04, p < .001, MSE_2 = 0.613$, as this represents the semantic interference manipulation. Sentences 6–9 were analyzed with a 2 (embedding type) × 2 (NP) within-subject ANOVA. An effect of embedding was found, suggesting that the adverbial phrase fit more naturally with the embedded clause from the LoSyn sentences than with the embedded clause from the HiSyn sentences, $F_1(1, 25) = 29.60, p < .001, MSE_1 = 0.150; F_2(1, 35) = 15.17, p < .001, MSE_2 = 0.404$. There was no difference associated with the NPs in the two sentence types (Fs < 1). There was a hint of an interaction in the analysis by subjects, such that the type of NP affected plausibility of the embedded clauses from the LoSyn conditions more than that in the HiSyn conditions, $F_1(1, 25) = 3.72, p = .07, MSE_1 = 0.023$, but this effect did not near significance in the analysis by items ($F_2 < 1$).

As mentioned previously, the materials from Experiment 2 were also modified so that the critical verb was followed by substantially more material (usually two prepositional phrases) so that a larger spillover and final region could be analyzed. This was done to ensure that these regions would be adequately fixated so that any effects that occurred after the critical region could be clearly measured. The regions for analysis are illustrated in Table 11.

Presentation of items in the actual experiment followed the procedure used in Experiment 2, with each participant taking part in each condition but receiving only one of the four conditions from each item set. Items were presented randomly, mixed with 108 filler items from unrelated experiments, totaling 144 sentences in the experiment.

Every experimental sentence was followed by a comprehension question presented in the same cloze format used in Experiment 2. As before, the question was followed by a set of choices three lines below it, and participants were instructed to press a button corresponding to their answer. Three choices were presented in the current experiment (compared with two in Experiment 2), reflecting each of the three NPs in the sentence that could plausibly fit with the critical verb (i.e., the same three NPs tested in Sentences 1–3 in the plausibility experiment; see Table 12).

Data Analysis

All dependent measures were analyzed using a 2 (syntactic interference) × 2 (semantic interference) ANOVA. In addition to accuracy on the comprehension question, an analysis of errors is presented. Online eye-tracking measures are reported on four regions of interest, indicated in Table 11. Analyses were conducted

Table 12
Sentences Submitted for Plausibility Judgments, Based on Experimental Materials for Experiment 3

Test sentence	Mean plausibility rating
1. The pilot moaned.	6.69
2. The lady moaned.	6.69
3. The man moaned.	6.72
4. The smelly seat moaned.	2.39
5. The smelly man moaned.	6.44
6. The lady was sitting in the smelly seat yesterday afternoon.	6.20
7. The lady was sitting near the smelly man yesterday afternoon.	6.10
8. The lady said that the seat was smelly yesterday afternoon.	5.73
9. The lady said that the man was smelly yesterday afternoon.	5.75

Note. Rating scale ranged from 1 (*not plausible*) to 7 (*highly plausible*).

as described in Experiment 2 for each dependent measure in each region. Condition means from the analyses with participants as the random factor are presented together with 95% CIs, calculated as described above. Across all reading time measures, trimming affected 2.6% of the data.

Results

Comprehension Questions

Accuracy. Condition means are presented in Table 13, and the results of ANOVA testing are presented in Table 14. The main effect of syntactic interference was observed (.81 for LoSyn vs. .71 for HiSyn; CI = .04), as was the main effect of semantic interference (.81 for LoSem vs. .71 for HiSem; CI = .04). The interaction was not significant; however, as predicted by the retrieval account, the semantic interference effect was present in both the LoSyn conditions, $F_1(1, 39) = 7.19, p < .02, MSE_1 = 0.034$; $F_2(1, 35) = 6.50, p < .02, MSE_2 = 0.036$, and the HiSyn conditions, $F_1(1, 39) = 19.02, p < .001, MSE_1 = 0.026$; $F_2(1, 35) = 11.01, p < .003, MSE_2 = 0.042$.

Errors. Incorrect answers to the comprehension question occur when participants choose either the matrix subject of the sentence or the distracting NP (e.g., *pilot* or *man* from the example in Table 11). This occurred in 24% of trials. Table 15 presents a frequency tally for each of these choices by condition. An overall chi-square statistic on this table was significant, $\chi^2(3) = 39.35, p < .001$. Separate analyses for the subject NP, which was identical in all conditions, revealed only a marginal difference in the distribution across conditions, $\chi^2(3) = 7.38, p = .06$. The effect of condition on the probability of choosing the distracting NP was much greater, $\chi^2(3) = 64.64, p < .001$. In particular, participants were almost twice as likely to choose the distracting NP when it was a syntactic subject (HiSyn/HiSem condition) than when it was not (LoSyn/HiSem) (68 vs. 36).[2]

Reading Time Measures

Table 16 summarizes the results for each measure in each region of interest. All analyses were conducted on raw reading times for accurate trials only.

Pre–critical region. Table 17 presents the results of ANOVA testing in this region. Only measures with significant effects are displayed. There were no significant effects in the first-pass reading times for the inserted adverbial phrase (*Fs* < 1). The syntactic manipulation was significant in the regression path measure (483 ms for LoSyn vs. 520 ms for HiSyn; CI = 29). Pairwise comparisons showed the effect to be nonsignificant for the LoSem conditions (*Fs* < 1.02). The HiSem conditions showed a significant effect in the analysis by items, $F_2(1, 35) = 7.98, p < .01, MSE_2 = 42{,}110$, but only a marginal effect in the analysis by participants, $F_1(1, 39) = 3.76, p < .06, MSE_2 = 35{,}638$. The semantic manipulation was significant in the analysis by participants (481 ms for LoSyn vs. 522 ms for HiSyn; CI = 37), but there was only a trend for an effect in the analysis by items ($p = .07$). The interaction was significant in the analysis by items but not in the analysis by participants. In pairwise comparisons, the semantic manipulation was not significant in the LoSyn conditions (472 ms for LoSem vs. 493 ms for HiSem; CI = 32, *Fs* < 1) but was marginal for the HiSyn conditions in the analysis by participants (490 ms for LoSem vs. 551 ms for HiSem; CI = 68), $F_1(1, 39) = 3.11, p < .09, MSE_1 = 47{,}626$, and significant in the analysis by items (492 ms for LoSem vs. 577 ms for HiSem; CI = 71), $F_2(1, 35) = 5.30, p < .03, MSE_2 = 51{,}460$.

For the total reading time measure, the effect of the syntactic manipulation was observed (613 ms for LoSyn vs. 666 ms for HiSyn; CI = 38). The effect of the semantic manipulation was also observed in this region (604 ms for LoSem vs. 675 ms for HiSem; CI = 43). The interaction was not significant in the analysis by participants but was significant in the analysis by items. Pairwise comparisons revealed that the semantic effect was greater in the HiSyn conditions (88 ms), $F_1(1, 39) = 5.79, p < .02$; $F_2(1, 35) = 12.14, p < .002$, than in the LoSyn conditions (54 ms), where it was marginal in the analysis with participants as the random factor, $F_1(1, 39) = 3.46, p < .07$, and nonsignificant in the analysis by items, $F_2(1, 35) = 2.14$.

Participants made regressions out of this region in 5% of trials. The syntactic manipulation was not significant in these data. The semantic manipulation was significant (.03 for LoSem vs. .06 for HiSem; CI = .026). The interaction was not significant.

Critical region. Table 18 displays the results of ANOVA testing in this region; only measures with significant effects are displayed. The results in this region are similar to those in Experiment 2, in that neither the semantic interference effect nor the interaction was significant for any measure. In the first-pass reading times, the effect of syntactic interference was marginal in the analysis by participants ($p = .06$) but significant in the analysis by items (280 ms for LoSem vs. 291 ms for HiSem; CI = 10). There was a marginal interaction ($p < .08$) in the analysis by participants, but this effect was not significant in the analysis by items. This was due to an effect of syntactic interference in the LoSem conditions (274 ms for LoSyn vs. 294 ms for HiSyn; CI = 16), $F_1(1, 39) = 5.86, p < .02, MSE_1 = 2{,}748$, replicating the effect originally observed by Van Dyke and Lewis (2003). The effect was only marginal in the analysis by items (277 ms for LoSyn vs. 295 ms for HiSyn; CI = 19.6), $F_2(1, 35) = 3.08, p < .09, MSE_2 = 3{,}934$. The syntactic effect was not significant in the HiSem conditions (*Fs* < 1).

Table 13
Accuracy Scores for Comprehension Questions in Experiment 3, With Participants as the Random Factor

Interference type	Accuracy
LoSyn/LoSem	.85 (.03)
LoSyn/HiSem	.77 (.03)
HiSyn/LoSem	.77 (.03)
HiSyn/HiSem	.66 (.03)

Note. Values in parentheses are standard errors. LoSyn and HiSyn refer to low and high syntactic interference conditions, respectively; LoSem and HiSem refer to low and high semantic interference conditions, respectively.

[2] The frequency of choosing the distractor in the LoSem conditions is not interesting, as the distracting NP was not present in the sentence. Those data show merely that when they are incorrect, participants prefer to choose an NP that occurred in the sentence they read instead of one that did not.

Table 14
Experiment 3: Analysis of Variance Results for Comprehension Questions

Measure	Main effect: Syntactic interference	Main effect: Semantic interference	Interaction
Accuracy	$F_1(1, 39) = 19.23, p < .001, MSE_1 = 0.019$ $F_2(1, 35) = 11.76, p < .003, MSE_2 = 0.029$ min$F'(1, 68) = 7.30, p < .01$	$F_1(1, 39) = 26.76, p < .001, MSE_1 = 0.013$ $F_2(1, 35) = 16.73, p < .001, MSE_2 = 0.020$ min$F'(1, 68) = 10.29, p < .005$	$F_1 < 1.00$, *ns* $F_2 < 2.22$, *ns*

For the regression path measure, the effect of syntactic interference was significant (315 ms for LoSyn vs. 349 ms for HiSyn; CI = 28). As before, this effect was more strongly attested in the LoSem conditions (314 ms for LoSyn vs. 354 ms for HiSyn; CI = 37), $F_1(1, 39) = 4.54, p < .04, MSE_1 = 14{,}016$; $F_2(1, 35) = 5.08, p < .04, MSE_2 = 7{,}783$. The effect was not significant in the HiSem conditions.

In the measure of total time in the region, there was a marginal effect of syntactic interference (418 ms for LoSyn vs. 459 ms for HiSyn; CI = 43, $p < .07$), but pairwise comparisons were not significant. Participants made regressions from this region in 9% of trials, but no significant effects were observed in these data.

Spillover region. No significant effects were found in the first-pass reading times. In the regression path measure, there were also no significant main effects observed (Fs < 2.79); however, pairwise comparisons revealed an effect of semantic interference in the LoSyn conditions, $F_1(1, 39) = 5.91, p < .02$, which was not significant in the analysis by items ($F_2 < 2.30$). This effect was caused by a *decrease* in reading times when the distracting NP was a suitable subject for the critical verb. The interaction was not significant.

No significant effects were observed in either the total reading time or the proportion of regressions. Participants made backward regressions out of this region in 14% of trials.

Final region. Table 19 shows the results of ANOVA testing in this region. Only measures with significant effects are displayed. No significant effects were observed in the first-pass reading times in this region (Fs < 1.07). For the regression path measure, the effect of syntactic interference was not significant. The effect of semantic interference was observed (1,972 ms for LoSyn vs. 2,310 ms for HiSyn; CI = 220). Pairwise comparisons revealed that the effect was significant in the HiSyn conditions (406 ms), $F_1(1, 39) = 5.60, p < .03$; $F_2(1, 35) = 4.25, p < .05$, and marginal in the LoSyn conditions (271 ms) in the analysis by participants, $F_1(1, 39) = 3.49, p = .07$, but nonsignificant in the analysis by items. This result is not consistent with the local coherence account, which predicts no effect for either construction, because local coherence is broken by the intervening adverbial phrase.

Table 15
Frequency of Choosing the Matrix Subject or the Distracting Noun Phrase (Pilot vs. Man From Table 11)

Interference type	Subject	Distractor
LoSyn/LoSem	41	10
LoSyn/HiSem	44	36
HiSyn/LoSem	66	15
HiSyn/HiSem	52	68

Note. LoSyn and HiSyn refer to low and high syntactic interference conditions, respectively; LoSem and HiSem refer to low and high semantic interference conditions, respectively.

There were no significant effects in the total reading time measure in this region (Fs < 1.54). Participants made regressions from the final region in 52% of trials, but no significant effects were observed (Fs < 2.45).

Discussion

This experiment provides support for the retrieval account of syntactic interference over a "stumbling" account, as the effect was observed in both first-pass reading times and regression path measures, despite the separation of the two verbs. It is particularly noteworthy that the contrast of the syntactic manipulation in the LoSem conditions was significant, as this comparison avoids any additional slowdown that could be caused by the semantic manipulation. Moreover, it is this pair that replicates the original demonstration of syntactic interference by Van Dyke and Lewis (2003).

The current evidence also suggests that the effect observed at the critical verb is not spillover of difficulty associated with processing the additional embedded clause in the HiSyn constructions, because that difficulty would have to extend all the way past the adverbial phrase for it to be observed at the critical verb. Although this possibility cannot be ruled out, it seems unlikely. One account of how this difficulty could arise depends on the additional storage cost associated with maintaining the syntactic prediction of the embedded clause that follows *declared* (Gibson, 2000; Grodner & Gibson, 2005). Figure 1 illustrates the predictions of dependency locality theory (DLT) in detail (see also Van Dyke & Lewis, 2003, for a similar discussion of these structures). The theory suggests that three types of processing costs determine resource expenditure and, hence, processing difficulty: (a) those associated with creating discourse referents; (b) those associated with integrating grammatical heads and dependents; and (c) those associated with storing incomplete structural dependencies. Figure 1 shows clearly that the HiSyn conditions are predicted to be more difficult than the LoSyn conditions in the embedded region, because three predictions must be maintained at the verb *said*: that for the verb that will eventually be filled by *moaned* and the two predictions for the subject and verb signaled by the complement-taking verb *said*, which will eventually be filled by *the seat was smelly*. This is in contrast to the single prediction

Table 16
Experiment 3: Raw Reading Times (in Milliseconds) and Proportion of Regressive Eye Movements for Each Region for Each Dependent Measure, With Participants as the Random Factor

Measure and interference type	Pre-critical region	Critical region	Spillover region	Final region
First pass				
LoSyn/LoSem	449 (15)	274 (9)	448 (24)	464 (26)
LoSyn/HiSem	455 (16)	282 (8)	437 (19)	480 (23)
HiSyn/LoSem	462 (18)	294 (11)	442 (21)	483 (29)
HiSyn/HiSem	447 (17)	280 (11)	424 (23)	463 (24)
Regression path				
LoSyn/LoSem	472 (17)	314 (14)	607 (33)	1,875 (200)
LoSyn/HiSem	493 (22)	315 (13)	553 (28)	2,147 (223)
HiSyn/LoSem	490 (21)	354 (18)	625 (30)	2,068 (232)
HiSyn/HiSem	551 (36)	344 (23)	615 (41)	2,474 (317)
Total time				
LoSyn/LoSem	586 (29)	414 (25)	711 (43)	649 (46)
LoSyn/HiSem	640 (41)	421 (25)	707 (47)	705 (43)
HiSyn/LoSem	622 (31)	451 (28)	731 (51)	675 (44)
HiSyn/HiSem	710 (47)	467 (38)	766 (59)	663 (46)
Proportion of regressions				
LoSyn/LoSem	.03 (.01)	.08 (.02)	.17 (.03)	.50 (.04)
LoSyn/HiSem	.04 (.01)	.08 (.02)	.11 (.02)	.54 (.04)
HiSyn/LoSem	.03 (.01)	.10 (.02)	.17 (.02)	.50 (.05)
HiSyn/HiSem	.08 (.02)	.13 (.03)	.16 (.02)	.52 (.05)

Note. Values in parentheses are standard errors. LoSyn and HiSyn refer to low and high syntactic interference conditions, respectively; LoSem and HiSem refer to low and high semantic interference conditions, respectively.

maintained at *was sitting*, which is waiting only for the verb *moaned*. The critical point is that this difference is resolved at the final word in this region (*seat* in the LoSyn conditions and *smelly* in the HiSyn conditions). After both the subject and the verb of the embedded clause have been processed, the storage cost associated with both sentences is back to one energy unit, associated with the still outstanding prediction for the verb *moaned*. From this point on (i.e., the adverbial phrase and the critical region), the processing cost is determined by the number of discourse referents that must be crossed in making the necessary integrations. The structures in the LoSyn condition actually have one additional discourse referent as compared with those in the HiSyn condition, and so the LoSyn conditions are predicted to be *more* difficult both in the adverbial phrase and in the critical region—the opposite of the pattern observed in the data.[3]

There remains the possibility that the effect observed in the critical region is due to spillover from processing the adverbial phrase that occurred in the pre-critical region, as there was a significant effect of the syntactic manipulation observed in that region in the regression path and the total time measure. The effect in the total time measure does not support the spillover explanation, because the measure includes fixations that may have originated anywhere in the sentence (even after the critical retrieval region). The effect in the regression path measure may support the spillover account. However, this explanation suggests that this effect engendered such a strong processing load that it influenced the first-pass times in the following (critical) region. This is not likely, because the pairwise comparison of the syntactic manipulation in the regression path measure in the pre-critical region for the LoSem conditions, which are the most direct test of the effect, was not significant.

Although the effects observed in the pre-critical region are associated with the interference manipulation, they cannot be due to retrieval interference per se because they occur at the adverbial phrase, which appears prior to the retrieval probe. Rather, it is likely that they reflect the significant difference in plausibility ratings for the embedded clause: The adverbial phrase was judged to be more plausible in the LoSyn conditions than in the HiSyn conditions, perhaps giving rise to longer backward regressions in the HiSyn conditions.

Further support for the interference account of the difference between the LoSyn and HiSyn conditions is evident in the analysis of errors to the comprehension question. Because participants were twice as likely to choose the distracting NP as the subject of the critical verb when it had subject case marking compared with when it did not, it appears that participants do entertain the incorrect grammatical relationship despite semantic evidence to the contrary. Neither the adjacency nor the embedding account makes this prediction regarding the error data.

Experiment 3 also provided further evidence for the semantic interference effect during online reading, and the results were consistent with previous experiments where the effect occurred downstream from the critical verb. The effect was observed in the regression path measure in the final region and, as in the two previous experiments, was particularly strong for the HiSyn conditions. The general pattern of the semantic interference manipu-

[3] I thank Ted Gibson for pointing out that reading times in the critical region for the HiSyn conditions could be increased if greater weight were given to tensed verbs than to other elements (i.e., prepositions, nouns, adjectives) in counting discourse units. Experiments to explore this hypothesis are currently underway.

Table 17
Experiment 3: Analysis of Variance Results for Reading Time Measures in the Pre–Critical Region

Measure	Main effect: Syntactic interference	Main effect: Semantic interference	Interaction
Regression path	$F_1(1, 39) = 6.59, p < .02, MSE_1 = 8{,}702$	$F_1(1, 39) = 4.89, p < .04, MSE_1 = 13{,}624$	$F_1 < 1.05$, *ns*
	$F_2(1, 35) = 6.23, p < .02, MSE_2 = 19{,}461$	$F_2 < 3.58$, *ns*	$F_2(1, 35) = 5.31, p < .03, MSE_2 = 10{,}110$
	min$F'(1, 57) = 3.20, p < .08$		
Total time	$F_1(1, 39) = 7.29, p < .01, MSE_1 = 15{,}110$	$F_1(1, 39) = 10.73, p < .003, MSE_1 = 18{,}672$	$F_1 < 1.00$, *ns*
	$F_2(1, 35) = 5.48, p < .03, MSE_2 = 27{,}769$	$F_2(1, 35) = 11.21, p < .003, MSE_2 = 20{,}039$	$F_2(1,35) = 4.33, p < .05, MSE_2 = 13{,}603$
	min$F'(1,71) = 3.13, p < .09$	min$F'(1,74) = 5.48, p < .03$	
Regressive eye movements	$F_1 < 3.13$, *ns*	$F_1(1, 39) = 4.78, p < .04, MSE_1 = 0.74$	$F_1 < 2.00$, *ns*
	$F_2 < 2.82$, *ns*	$F_2(1, 35) = 4.15, p < .05, MSE_2 = 1.00$	$F_2 < 1.00$, *ns*
		min$F'(1, 73) = 2.22$, *ns*	

lation occurring earlier for the LoSyn conditions did not hold up, probably because of the larger regions defined in this study as compared with Experiments 1 and 2. In those experiments, the effect in the LoSyn conditions occurred in the region containing the two words prior to the final word, which were included in the final region defined for the current experiment. Whereas there was no evidence for a semantic interference effect at the final word in the LoSyn conditions in the previous experiments, the marginal effect observed in the final region in Experiment 3 is probably the same effect seen there.

The pattern of early syntactic effects and later semantic effects is generally consistent with a number of other studies. The appearance of the syntactic interference effect in the first-pass measure is consistent with the view that this measure reflects the ease of structural integration (Boland, 2004; Boland & Blodgett, 2001; Frazier & Rayner, 1982) and the results of Van Dyke and Lewis (2003) suggesting that interference has its effect on the creation of grammatical dependencies. The appearance of the semantic interference effect in the final region of the online measures is consistent with other studies that have shown effects of discourse or semantic anomaly later than the region containing the anomaly (Boland & Blodgett, 2001; Braze, Shankweiler, Ni, & Palumbo, 2002; Ni, Fodor, Crain, & Shankweiler, 1998). I return to a discussion of the time course of these effects in the General Discussion. Experiment 3 also provides more evidence for strong syntactic and semantic interference effects in the offline comprehension measure. It appears that although the interference effects may be noticed during online reading, either in the critical region for the syntactic manipulation or in the final region for the semantic manipulation, they are difficult to resolve. The fact that the effects observed here occurred even in the absence of adjacency of either the embedded verb with the critical verb or the semantic distractor with the critical verb challenges accounts associated with local coherence—either the lack of coherence due to two sequential verbs or the momentary coherence due to the semantic fit of the distractor and the critical verb. The retrieval view provides a more adequate account of the data because it predicts difficulty whenever an NP that partially matches the retrieval probe occurs at any intervening position between the probe and the target.

One aspect of the data that was unexpected is the significant decrease in regression path reading times for the LoSyn/HiSem condition in the spillover region. It is evident from the comprehension questions and the reading times at the final region that participants do find this condition more difficult than its control (LoSyn/LoSem) condition, so this facilitation is a temporary effect. Further experimentation will be necessary to verify the generality of this effect.

General Discussion

The research program motivating the experiments presented here is to make explicit the link between memory processes and language comprehension. Memory interference is a basic finding in the memory literature and is understood as a primary constraint on accessing stored information (e.g., Anderson & Neely, 1996; Crowder, 1976; Nairne, 2002). A number of studies have shown reduction in recognition and recall of an item when it is semanti-

Table 18
Experiment 3: Analysis of Variance Results for Reading Time Measures in the Critical Region

Measure	Main effect: Syntactic interference	Main effect: Semantic interference	Interaction
First pass	$F_1 < 3.96$, *ns*	$F_1 < 1.00$, *ns*	$F_1 < 3.35$, *ns*
	$F_2(1, 39) = 4.28, p < .05, MSE_2 = 927$	$F_2 < 1.00$, *ns*	$F_2 < 1.05$, *ns*
Regression path	$F_1(1, 39) = 6.15, p < .02, MSE_1 = 7{,}792$	$F_1 < 1.00$, *ns*	$F_1 < 1.00$, *ns*
	$F_2(1, 35) = 9.34, p < .005, MSE_2 = 4{,}140$	$F_2 < 1.00$, *ns*	$F_2 < 1.00$, *ns*
	min$F'(1, 72) = 3.71, p < .06$		

Table 19
Experiment 3: Analysis of Variance Results for Reading Time Measures in the Final Region

Measure	Main effect: Syntactic interference	Main effect: Semantic interference	Interaction
Final region: Regression path	$F_1 < 2.79$, *ns* $F_2 < 3.58$, *ns*	$F_1(1, 39) = 9.26$, $p < .005$, $MSE_1 = 494{,}089$ $F_2(1, 35) = 6.42$, $p < .02$, $MSE_2 = 741{,}409$ min$F'(1, 70) = 3.79$, $p < .06$	$F_1 < 1.00$, *ns* $F_2 < 1.00$, *ns*

cally similar to other studied items, in circumstances where the reduction cannot be attributed to decay (e.g., Gorfein & Jacobson, 1972, 1973; Petrusic & Dillon, 1972; Watkins & Watkins, 1975; Wickens, 1970). From this perspective, the demonstration of semantic interference presented here is not surprising, although it is quite unexpected from the point of view of those language processing models that give pride of place to syntactic representations. The current data suggest that NPs whose syntactic encoding should make them unavailable to serve as the subject of a verb are nevertheless recruited to do so when they are semantically suitable. This result converges with that of Tabor et al. (2004) and extends those findings to show that interference effects may arise in the absence of locally consistent interpretations.

Evidence from the memory literature supports a "cue overload" account of these types of effects, where incorrect retrievals occur because the cues available at retrieval do not uniquely distinguish the target item. Similar distractors are retrieved instead, on the basis of a partial feature match with the retrieval probe via Equation 1. The impact of this retrieval mechanism on sentence processing is that erroneous interpretations may arise if the match between a distractor and the grammatical head supplying the retrieval cues is sufficiently strong. For example, in the syntactic interference effect, the intervening subject NP *seat* in Sentence 9 is retrieved as the subject of the VP *moaned* despite there being no grammatical licensing for this NP to serve as the subject of both the clause headed by *was smelly* and that headed by *moaned*, and despite being inappropriate as the type of object that can *moan*.

(9) The pilot remembered that the lady who said that the seat was smelly moaned.

Similarly, in the semantic interference effect, the NP *man* in Sentence 10 is interpreted as the subject of the VP *moaned* on the basis of its semantic match with the retrieval cues, despite its previous assignment in a nonsubject syntactic role, which should provide evidence against such an interpretation.

	Input word							Embedded Region						Adverbial Phrase		Critical Region
Cost type	The	pilot	remembered	that	the	lady	who	was	sitting	in	the	smelly	seat	yesterday	afternoon	moaned...
New referent	0	1	1	0	0	1	0	0	1	1	0	1	1	1	1	1
Attachment	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	6
Storage	2	1	0	2	2	1	3	1	1	2	2	2	1	1	1	0
Total	2	2	1	2	2	2	3	1	2	3	2	3	2	6	2	7

Cost type	The	pilot	remembered	that	the	lady	who	said	that	the	seat	was	smelly	yesterday	afternoon	moaned...
New referent	0	1	1	0	0	1	0	1	0	0	1	0	1	1	1	1
Attachment	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	5
Storage	2	1	0	2	2	1	3	3	3	3	2	2	1	1	1	0
Total	2	2	1	2	2	2	3	4	3	3	3	2	2	5	2	6

Figure 1. Word-by-word predictions of the dependency locality theory for the LoSyn/LoSem (top panel) and HiSyn/LoSem (bottom panel) conditions. Numbers represent units of cost, with regions of higher cost associated with longer reading times and higher processing difficulty.

(10) The pilot remembered that the lady who was sitting near the smelly man moaned.

These effects present a challenge for grammar-driven parsers (e.g., Fodor & Frazier, 1980; Fodor & Inoue, 1998; Frazier, 1987; Frazier & Clifton, 1996; Gibson, 2000; Gorrell, 1995; Sturt & Crocker, 1996), which would not predict difficulties associated with constituents already integrated into the existing parse tree, regardless of any featural match with later occurring items. In fact, most parsers contain explicit right-edge or configurational constraints on attachment that would bar the interpretation of the interfering NP as the subject of the critical verb in these sentences. One parser in this class with a mechanism that may predict the interference effects in the LoSyn conditions is the *attach anyway model* of Fodor and Inoue (1994, 1998, 2000), which includes a mechanism called "theft." Theft allows adjacent items to be "stolen" at the word level, without respecting existing constituent structure, in order to attach an otherwise stranded item into the current parse. It is intended as a method of "jumpstarting" reanalysis, by creating an inconsistency that will provoke assignment of new argument structure and thereby uncover the globally correct attachment site for the previously stranded item. The problem that the interference effect in the LoSyn constructions presents for this mechanism is that these effects occur in the absence of grammatical ambiguity; the item does actually have an open attachment site in the parse tree, but this memory representation has become inaccessible and so attachment fails. In addition, the data from Experiment 3 demonstrate that adjacency is not a necessary condition for these ungrammatical attachments to occur. Rather, it seems that a model like Fodor and Inoue's would need to become explicit about how the parser interacts with underlying memory mechanisms in order to account for these effects. In addition, adapting parsers in the class to account for the semantic interference effects is likely to be even more difficult, because these parsers are typically blind to the semantic properties of the lexical items contained in a sentence.

One possibility for understanding the relationship between the parser and the memory system is to assume that each word in a sentence enters memory as a syntactically elaborated feature structure. This creates a memory set analogous to those presented in short-term memory experiments, except that syntactic properties of the items describe interrelationships among the list items. It is important to note that it is these syntactic properties that initiate retrievals whenever a grammatical dependency must be created. Specifically, the feature structure of the head of a grammatical dependency, which is assumed to include information regarding the syntactic and semantic properties of its dependents (e.g., argument structure, subcategorization restrictions), provides the retrieval cues used to identify dependents in the sentential list. It should be noted that the assumption of a set of memory items is not incompatible with a structured representational form, such as a parse tree. The critical insight for understanding cue-based retrieval is that although a structured tree is necessary to describe syntactic relationships between items, the retrieval mechanism does not access items via the tree but rather does so directly, via cues. This type of access has been termed *direct access* or *content addressable access* and has been demonstrated in sentence processing by McElree (2000; McElree et al., 2003), who showed that the speed of processing structures with increasing distance between two grammatically dependent items, as well as increasing the structural complexity of the interpolated material, remained constant. Thus, retrieval speed is unaffected by the position of the retrieved item in the parse tree. This would not be the case if the parser executed a step-by-step traversal back through the parse tree to locate items available for completing dependencies, as the length of this kind of search would vary with the increased distance and/or complexity of the intervening material. Rather, direct access occurs by matching all items *in parallel* against the retrieval probe, regardless of any previous syntactic commitment.[4] The parallel match yields retrieval speeds that are unaffected by the number of items in memory and the relationships between them but also opens the door to retrieval interference from nontarget items that share only a subset of the retrieval cues, causing them to be retrieved inappropriately.

In contrast to grammar-based parsers, constraint-based parsers (e.g., MacDonald et al., 1994; Stevenson, 1994, 1998; Stevenson & Merlo, 1997; Vosse & Kempen, 2000) offer more promise for integrating the dynamical properties of direct access and the presence of interference effects because of their ability to access the full feature structure of items in a sentence. One model in this class that is particularly well suited to predict both the on- and the offline results observed here is the self-organizing parser (SOPARSE) model of Tabor and colleagues (Tabor, 2006; Tabor & Hutchins, 2004). In this model, sentence parses are built up via collections of lexically anchored tree fragments. Any given tree fragment will bond to every other fragment with which it has compatible features. Because the model does not contain explicit well-formedness constraints, some of these bonds will correspond to grammatically invalid relationships. To control the proliferation of ungrammatical bonds, the model relies on the principle of self-organization, wherein structures that are locally optimal with respect to the model's dynamics win out over those more weakly supported by evidence from the input. Most of the time the globally inconsistent bonds die off, but they may not, giving rise to a type of interference effect resembling those observed here. Thus, it is a distinguishing prediction of this model that interference effects should occur, as the model has no way to restrict interactions based on grammaticality, adjacency, or other configurational constraints that might rule them out.

From the point of view of a cue-based parser, the effects observed here have quite a different source, arising from general principles of the working memory system (Lewis et al., 2006). It is a basic fact of the memory system that the amount of information that can be actively maintained in memory is severely limited (see McElree, 2006, for a recent review). Changes in activation may cause linguistic representations to become weakened or inaccessible, causing the grammatical relationships that require them to be difficult—perhaps even impossible—to construct. At a behavioral level, this would give rise to the enduring offline effects

[4] The dynamics of this property have been well studied in models of episodic memory retrieval (e.g., Search of Associative Memory [SAM]: Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1981; Theory of Distributed Associative Memory [TODAM]: Murdock, 1982, 1983; Matrix: Pike, 1984; MINERVA2: Hintzman, 1984, 1988), where it has been referred to as "global matching." See S. E. Clark and Gronlund (1996) for a review.

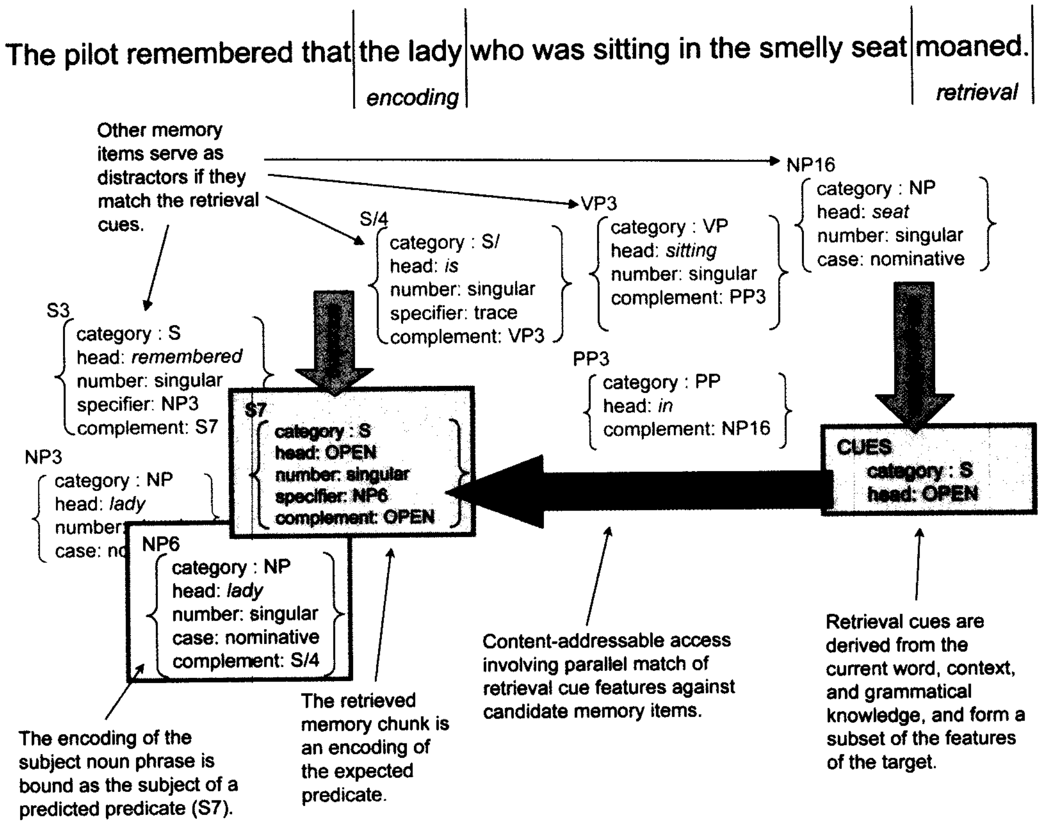

Figure 2. An illustration of the Lewis and Vasishth (2005) model parsing a LoSyn/LoSem sentence. S = subject; NP = noun phrase; VP = verb phrase; PP = prepositional phrase.

observed in the comprehension data. The incorrect interpretation that they *do* construct is the result of the parallelism in the system: Retrieval cues are matched simultaneously to all items in memory, and those with the highest match vis-à-vis Equation 1 have the highest probability of being restored into the focus of attention so that a grammatical dependency can be constructed.

At first blush, this highly limited capacity appears to place insurmountable constraints on language processing. Lewis and Vasishth (2005) offered one implementation of a cue-based parser that demonstrates that this is not the case. Their model was implemented within the ACT–R cognitive architecture, which incorporates the two defining features of cue-based retrieval—limited memory capacity and content addressability—in a way consistent with Equation 1.[5] The model's memory consists of chunks representing the syntactic structure built so far, together with predictions for constituents licensed by the current state of the parse. These chunks are not actively held in memory and decay as a function of time and prior retrievals. The only access to these items is via a retrieval buffer with the capacity to hold a single chunk. This affords the model the minimum capacity required to create new linguistic relations: the item waiting to be integrated into the parse and the chunk that licenses it. The item that is waiting is in the focus of attention and does not need to be retrieved. The chunk that licenses it is retrieved via the cues derived from the features of the waiting item.

Figure 2 illustrates how this process works to create the long-distance dependency in the LoSyn/LoSem constructions studied in the current experiments. The figure shows a collection of bracketed feature structures portraying the linguistic constituents stored in memory, each with an associated label (e.g., S3, NP16), which is used as an indexing structure. Those surrounded in gray boxes are pertinent to the current discussion; the darkened boxes are the only items currently active in memory. The parser works as follows: As each linguistic item is heard its syntactic configuration is encoded in chunks associated with each grammatical head. When the NP *lady* is heard, it is encoded as the head of an NP (NP6), and a prediction is created for a clause for which it can serve as the subject (S7). The parser then shifts focus in order to integrate the next words of the sentence, and these memory chunks decay as a function of their use in subsequent attachments. When the verb *moaned* occurs, it needs to both find its subject and find a way to

[5] Although the activation of an item in ACT–R's memory is determined by a slightly different equation, it is a function of both the strength of associations from retrieval cues and the target item (i.e., the numerator of Equation 1) and the number of items associated with the target (the denominator of Equation 1). Further details of the ACT–R architecture are available in Lewis and Vasishth (2005) and J. R. Anderson et al. (2004).

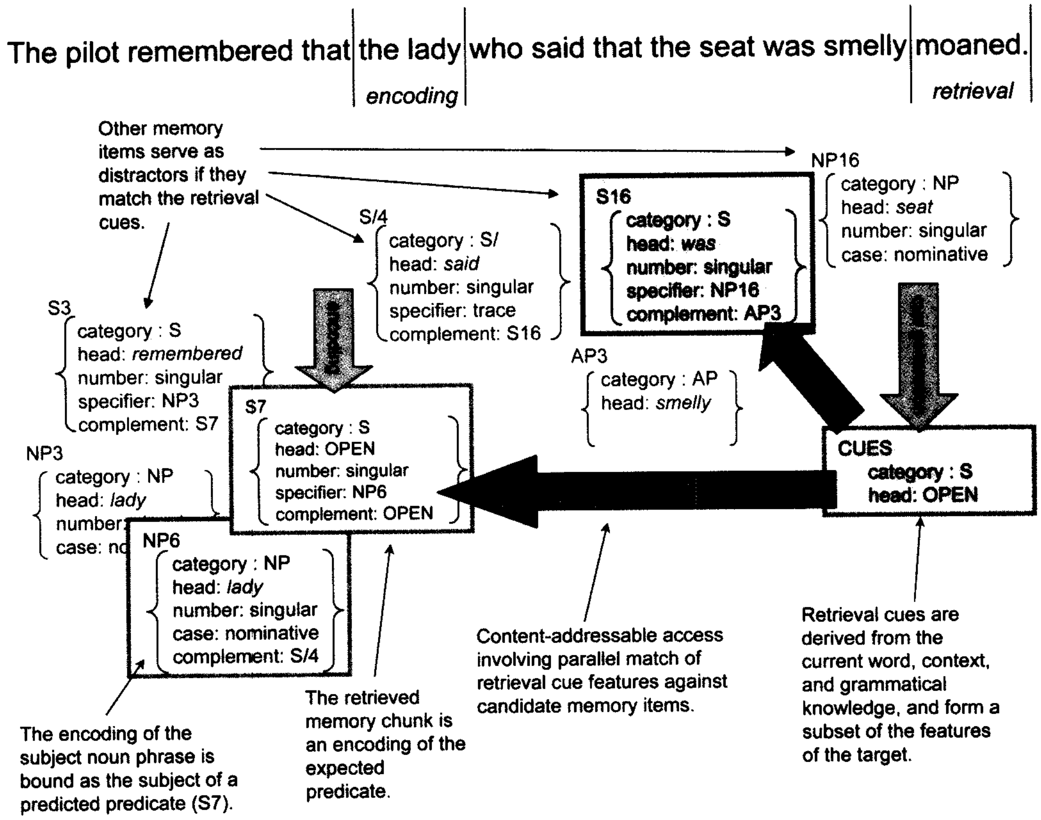

Figure 3. An illustration of the Lewis and Vasishth (2005) model parsing a HiSyn/LoSem sentence. S = subject; NP = noun phrase; VP = verb phrase; PP = prepositional phrase.

be integrated into the tree. The parser accomplishes this in a single step: setting retrieval cues for a predicted clause which will have been generated from a previously parsed NP. Thus, a long-distance dependency between the current word and one that occurred eight words prior to it was created, without the need to assume that the word was maintained in memory while the intervening words were processed. This makes the cue-based parsing account particularly desirable because it captures the architectural constraints that the memory system sets on language processing while still maintaining its impressive functionality.

Moreover, it is exactly this limited capacity architecture that gives rise to the interference effects demonstrated in the current experiments. Unlike the dependency locality theory discussed earlier, which predicts that the HiSyn construction will be more difficult because of the cost associated with *storing* an embedded syntactic prediction, the cue-based account predicts no cost for storing the embedded prediction per se because it is not necessary to maintain the prediction in active memory. Instead, the presence of the embedded prediction has its effect at retrieval, when the retrieval cues match to the embedded prediction instead of to the target prediction. This is illustrated in Figure 3, which depicts the Lewis and Vasishth (2005) parser processing a HiSyn/LoSem construction. As in Figure 2, a prediction for a clause (S7) is created when the NP *lady* (NP6) is encoded, and these items move out of active memory as the parser processes the intervening information. In this case, part of that intervening information is the construction of the embedded clause (S16), with *seat* (NP16) as its subject. When *moaned* is heard, it generates the same retrieval cues as in the previous example—for an S predicted by a prior NP. Retrieval then operates according to Equation 1, where each retrieval cue is matched in parallel to all of the items in memory. Memory items S3, S7, and S16 all have a partial match with the retrieval cues, but S16 will be the stronger match because it is more recent (i.e., its base activation is higher; see Lewis & Vasishth, 2005, for details of the ACT–R equations governing this match, and see footnote 5). This means the probability that the parser will bring S16 back into focus is much greater (and in particular, greater than that for S7, which is the correct licensor for *moaned*), leading to an incorrect and ungrammatical attachment.[6]

[6] The Lewis and Vasishth (2005) model follows the convention from head-driven phrase-structure grammar and categorical grammar (Pollard & Sag, 1994; Steedman, 1996) of using gapped, or *slash*, features. In Figure 3, S/4 is an example of one of these structures, and it will not match as strongly to the retrieval cues as the other clauses in memory because of the different category type (i.e., a slashed clause, shown as S/).

Although the Lewis and Vasishth (2005) model has not yet been tested on the semantic effects demonstrated here, this would be a worthwhile project for future work. Because it is the property of content addressability that gives rise to both the syntactic and the semantic interference effects, extending the model would require merely expanding the set of encoding features involved in the match to include the semantic (and perhaps referential) properties of the linguistic items. The model has already been shown to be robust at simulating the syntactic interference effect observed here and in Van Dyke and Lewis (2003). It should be emphasized, however, that the goal of this article is not to advocate a particular model of sentence processing, and indeed, the Lewis and Vasishth model is only one possible implementation of cue-based retrieval parsing. Rather, the goal is to demonstrate how properties of the memory retrieval system operate in language processing, giving rise to interference effects resembling those that have been well studied in other domains of memory research.

Although other explanations of syntactic complexity are present in the literature (e.g., Gibson, 1991, 1998; Just & Carpenter, 1992), these accounts mainly implicate storage as the source of processing difficulty. The semantic interference effect challenges these accounts because it suggests that specific features of the lexical items create difficulty during long-distance attachment, even when storage load is held constant. To account for these results, storage theories would have to incorporate additional assumptions about why storing items with particular semantic features creates difficulty whereas storing items with other features does not. In fact, the dependency locality theory has already moved in this direction (Warren & Gibson, 2002, 2005) by appealing to the notion of a referential hierarchy as modulating the cost of storing particular items; however, the theory does not yet include a mechanism for explaining the semantic interference results observed here. In contrast, the cue-based parsing framework gives a natural explanation for these effects that simultaneously reconciles the connection between the parsing mechanism and the memory system that supports it.

Time Course of Effects

Whereas the online effects of syntactic interference were observed in the region where the critical retrieval takes place, the effects of semantic interference were observed later, in regions that were identical across all conditions and contained no explicit prompt for retrieval. Consequently, I have argued that these effects are associated with reanalysis of incorrect grammatical dependencies produced by retrieving the semantically interfering noun as the subject of the critical verb. At issue, then, is why these effects are delayed past the point where the critical verb was incorrectly integrated into the sentence. Although the data in the current study do not provide a conclusive explanation for the apparent difference in time course, several potential sources can be noted. First, the results are consistent with those of McElree and Griffith (1995, 1998), who found that violations of syntactic constituent structure were noticed 50–100 ms before violations of theta roles (i.e., semantic fit) in a speed–accuracy trade-off paradigm. They interpreted this result as suggesting that syntactic processes have a faster rate of information accrual, perhaps owing to the more constrained nature of syntactic information (i.e., words are mapped onto a finite set of grammatical roles, which are combined according to a finite set of grammatical rules). In contrast, the evaluation of "semantic fit" between a verb and its argument is a less deterministic process, with varying degrees of fit possible. For example, interpreting *table* as the object of *began* in Sentence 11a leads to longer reading times as compared with when it is the object of *built* in Sentence 11b because of semantic properties of the verb, yet both interpretations are acceptable (McElree, Traxler, Pickering, Seely, & Jackendoff, 2001).

(11a) The carpenter began the table during the morning break.

(11b) The carpenter built the table during the morning break.

Thus, one explanation for the later occurring semantic interference effect is that it simply takes longer for individuals to compute the semantic association between the distractor and the verb and hence longer for the inconsistent assignment of one NP in two thematic roles to be recognized.

An alternative account for the timing difference focuses on the fact that the semantic interference effect occurs at the end of the sentence and may be part of sentence wrap-up processing. Although these effects have long been acknowledged, no clear account of exactly what occurs during this wrap-up processing has been proposed. One possibility stems from the coordination between a "fast and frugal" heuristic system with the output of more rigorous syntactic algorithms, as recently proposed by a number of researchers (Bever & Townsend, 2001; Ferreira, 2003). It is possible that the retrieval mechanism proposes candidate grammatical dependencies during online processing, which are then integrated into a consistent situation model during sentence wrap-up, using the full complement of linguistic knowledge to determine whether they are well formed (i.e., syntactic rules together with semantic, discourse, and pragmatic knowledge). In this case, online semantic interference effects may not be predicted in the critical region, as inconsistencies will become apparent only when the full set of proposed grammatical dependencies is considered.

Good Enough 'Cause That's All You've Got

Although further research is required to understand the mechanism underlying these time course differences, the data from all three experiments clearly suggest that in a large proportion of cases (25% in Experiment 3; 17% in Experiments 1 and 2), participants are unable to successfully reanalyze incorrect interpretations. This notion is consistent with a true interference effect: Distracting items block access to target items, making the targets unavailable for retrieval and hence unavailable for both initial attachment and reanalysis (cf. Van Dyke & Lewis, 2003, for a discussion of the relationship between these operations). Thus, in a HiSyn/HiSem sentence like 12, the retrieval account predicts that the retrieval cues from the verb *was complaining* are simply insufficient to retrieve the proper subject, *resident*, because of its high featural overlap with both syntactic and semantic features of *neighbor*. It is quite possible that readers realize that their initial interpretation of *neighbor* as the subject of *was complaining* violates grammatical constraints but are left no other choice because *neighbor* and *resident* (and *worker*) are indistinguishable vis-à-vis the retrieval cues from *was complaining*. This explanation would account for the strikingly low comprehension rates for these conditions (.65 for Experiment 3) despite significant effort in rereading (771 ms longer duration for regression paths from the final region as compared with the control condition).

(12) The worker was surprised that the resident who said that the neighbor was dangerous was complaining about the investigation.

This situation is akin to what Ferreira and colleagues have termed "good enough" language processing (Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Ferreira, Bailey, & Ferraro, 2002; Ferreira, Christianson, & Hollingworth, 2001), in which ambiguous constructions such as that in Sentence 13 cause participants to answer "yes" to both the question "Did the baby play in the crib?" and the question "Did Anna dress the baby?"

(13) While Anna dressed the baby played in the crib.

Although the interpretation of Anna dressing the baby is ultimately incorrect, there is a temporary ambiguity prior to the occurrence of *played*, during which this interpretation is valid. Ferreira and colleagues suggested that participants' incorrect responses to "Did Anna dress the baby?" come from this left-over representation, which has not been completely reanalyzed. They suggested that this arises from incomplete dismantling of structure that occurs after the theft mechanism (Fodor & Inoue, 1994, 1998, 2000) has "stolen" the adjacent NP as the subject of *played* (Christianson et al., 2001; Ferreira et al., 2001). The current results suggest that this is an incomplete explanation because adjacency is not required to produce misinterpretations such as those observed both by Ferreira et al. and in the current experiments.

Cue-based parsing provides a straightforward account of the mechanism producing these results, based on semantic interference and the lack of necessary retrieval cues for reanalysis. For example, in Sentence 13, the verb *played* supplies two retrieval cues: that for a clause for it to head and that for a subject. The former cues are satisfied by the adjunct clause built so far (*While Anna dressed the baby*), which expects to modify a main clause, and so *played* is integrated as the matrix clause of this construction. For the subject, the cues from *played* match both *baby* and *Anna* because of semantic overlap. Although *baby* is an imperfect match because it lacks subject case marking, it will have a higher probability of retrieval than *Anna* because it is more recent, and it will be retrieved as the subject of *playing*. This results in an ill-formed structure, where *baby* is encoded as both the subject of *played* and the object of *dressed*. The only way this structure could be made acceptable is if the alternative lexical frame of the verb *dressed* (as in *Anna dressed herself*) were substituted for the transitive frame, but *played* provides no retrieval cues for accessing it. The parser is stuck with an inconsistent representation, and readers are left no choice but to muddle on (cf. Fodor & Inoue, 1998, for a similar account).

What is particularly noteworthy about the current study is that it demonstrates cases where the parser gets stuck in the absence of ambiguity and when there is no local coherence to mislead. Moreover, the Christianson et al. (2001) results have been criticized because of the possibility that merely presenting the alternative question "Did Anna dress the baby?" was sufficient to activate the incorrect interpretation (Tabor et al., 2004). This interpretation is less likely for the current results, because participants were presented with a cloze comprehension question and had only to indicate which NP they thought was the subject of the critical verb. Thus, the cue-based parser provides a unified account of ungrammatical attachments that arise during the processing of both ambiguous and unambiguous constructions.

Conclusion

A popular assumption in psycholinguistics is that readers and listeners "do not violate their knowledge of grammar in arriving at an interpretation of a sentence" (Frazier & Clifton, 1996, p. 3). I have suggested that interference effects can sometimes leave readers with no choice but to do so. I have argued that these effects are due to properties of the memory retrieval mechanism, which supplies constituents for participation in grammatical dependencies. Because all items previously stored in memory are matched in parallel to the cues provided by the grammatical head, items that either partially or wholly overlap features of the target have a probability of being retrieved that is proportional to that overlap. This means that linguistic constituents whose previous syntactic commitments should make them unavailable for participating in a new grammatical dependency may nevertheless be retrieved to participate in new grammatical dependencies. In some cases, this may initiate successful reanalysis of incorrect attachments, and in others it may result in attachments that are simply ungrammatical, leading to erroneous interpretation.

Although the semantic interference effects are unexpected from the point of view of many parsing theories, they are clearly predicted by a long history of interference effects in the memory literature. From this perspective, the syntactic interference effects observed here are novel (although perhaps not so from a linguistic perspective). They suggest that syntactic codes are utilized during both encoding and retrieval of linguistic information and represent a dimension along which interference effects can arise (Lewis, 1996, 1999). Given the pervasiveness of interference effects in studies of memory capacity and forgetting, demonstrating the existence of these effects in sentence processing is crucial to any theory that seeks a close relationship between the mechanisms for language processing and the more general mechanisms of memory.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of mind. *Psychological Review, 111*, 1036–1060.

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork & R. A. Bjork (Eds.), *Memory (Handbook of perception and cognition)* (2nd ed., pp. 237–313). San Diego, CA: Academic Press.

Bever, T., & Townsend, D. J. (2001). *Sentence comprehension: The integration of habits and rules.* Cambridge, MA: MIT Press.

Boland, J. E. (2004). Linking eye movements to sentence comprehension in reading and listening. In M. Carreiras & J. C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond* (pp. 51–76). Brighton, England: Psychology Press.

Boland, J. E., & Blodgett, A. (2001). Understanding the constraints on syntactic generation: Lexical bias and discourse congruency effects on eye movements. *Journal of Memory and Language, 45*, 381–411.

Braze, D., Shankweiler, D., Ni, W., & Palumbo, L. (2002). Readers' eye movements distinguish anomalies of form and content. *Journal of Psycholinguistic Research, 31,* 25–44.

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

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

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

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

Crowder, R. G. (1976). *Principles of learning and memory.* Hillsdale, NJ: Erlbaum.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47,* 164–203.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*(1), 11–15.

Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research, 30,* 3–20.

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

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

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

Fodor, J. D., & Inoue, A. (1998). Attach anyway. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 101–142). Boston: 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 perspectives on language processing* (pp. 21–61). Boston: Kluwer.

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

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

Frazier, L. (1990). Exploring the architecture of the language system. In G. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 409–433). Cambridge, MA: MIT Press.

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

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

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14,* 178–210.

Gibson, E. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown* (Unpublished doctoral dissertation, Center for Machine Translation Technical Report No. 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.

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,* 1411–1423.

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language, 51,* 97–114.

Gorfein, D. S., & Jacobson, D. E. (1972). Proactive effects in short-term recognition memory. *Journal of Experimental Psychology, 95,* 211–214.

Gorfein, D. S., & Jacobson, D. E. (1973). Memory search in a Brown–Peterson short-term memory paradigm. *Journal of Experimental Psychology, 99,* 82–87.

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

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science, 29,* 261–290.

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

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

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,* 122–149.

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

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

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

Lewis, R. L. (1999). Accounting for the fine structure of syntactic working memory: Similarity-based interference as a unifying principle. *Behavioral and Brain Sciences, 22,* 105–106.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29,* 375–419.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences, 10,* 447–454.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1,* 476–490.

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

Markwardt, F. C. J. (1998). *Peabody Individual Achievement Test—Revised (PIAT–R).* Circle Pines, MN: American Guidance Service.

Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology, 5,* 203–220.

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

McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 46). San Diego, CA: Academic Press.

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

McElree, B., & Griffith, T. (1995). Syntactic and thematic processing in sentence comprehension: Evidence for a temporal dissociation. *Journal 2of Experimental Psychology: Learning, Memory, and Cognition, 21,* 134–157.

McElree, B., & Griffith, T. (1998). Structural and lexical constraints on filling gaps during sentence processing: A time-course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 432–460.

McElree, B., Traxler, M., Pickering, M. J., Seely, R., & Jackendoff, R. (2001). Reading time evidence for enriched composition. *Cognition, 78,* B17–B25.

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.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18,* 251–269.

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology, 53,* 53–81.

Ni, W., Fodor, J. D., Crain, S., & Shankweiler, D. (1998). Anomaly detection: Eye movement patterns. *Journal of Psycholinguistic Research, 27,* 515–539.

Petrusic, W. M., & Dillon, R. F. (1972). Proactive interference in short-term recognition and recall memory. *Journal of Experimental Psychology, 95,* 412–418.

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

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

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

Schneider, W. (1995). MEL Professional (Version 2.0) [Computer software]. Pittsburgh, PA: Psychology Software Tools.

Steedman, M. (1996). *Surface structure and interpretation.* Cambridge, MA: MIT Press.

Stevenson, S. (1994). Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research, 23,* 295–322.

Stevenson, S. (1998). Parsing as incremental restructuring. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 327–364). Boston: Kluwer.

Stevenson, S., & Merlo, P. (1997). Lexical structure and parsing complexity. *Language and Cognitive Process, 12,* 349–399.

Stolz, W. S. (1967). A study of the ability to decode grammatically novel sentences. *Journal of Verbal Learning and Verbal Behavior, 6,* 867–873.

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

Tabor, W. (2006, March). *A unified, self-organizing model of garden path phenomena, center-embedding phenomena, and interference effects.* Paper presented at the 19th Annual CUNY Conference on Human Sentence Processing, New York, NY.

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language, 50,* 355–370.

Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: Digging in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 431–450.

Van Dyke, J. A. (2002). *Retrieval effects in sentence parsing and interpretation.* Unpublished doctoral dissertation, University of Pittsburgh, Pittsburgh, PA.

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A retrieval interference theory of recovery from misanalyzed ambiguities. *Journal of Memory and Language, 49,* 285–316.

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language, 55,* 157–166.

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.

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

Warren, T., & Gibson, E. (2005). Effects of NP type in reading cleft sentences in English. *Language and Cognitive Processes, 20,* 751–767.

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory, 104,* 442–452.

Wickens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review, 77,* 1–15.

Received January 31, 2006
Revision received December 20, 2006
Accepted December 21, 2006 ■