Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Memory and Language 55 (2006) 157–166
Journal of Memory and Language
www.elsevier.com/locate/jml

Retrieval interference in sentence comprehension

Julie A. Van Dyke [a,b,*], Brian McElree [b]

[a] Haskins Laboratories, 300 George Street, New Haven, CT 06511-6695, USA
[b] New York University, USA

Received 8 December 2005; revision received 21 March 2006
Available online 9 June 2006

Abstract

The role of interference effects in sentence processing has recently begun to receive attention, however whether these effects arise during encoding or retrieval remains unclear. This paper draws on basic memory research to help distinguish these explanations and reports data from an experiment that manipulates the possibility for retrieval interference while holding encoding conditions constant. We found clear support for the principle of cue-overload, wherein cues available at retrieval cannot uniquely distinguish among competitors, thus giving rise to interference effects. We discuss the data in relation to a cue-based parsing framework (Van Dyke & Lewis, 2003) and other interference effects observed in sentence processing (e.g., Gordon, Hendrick, & Johnson, 2001, 2004). We conclude from the available data that the memory system that subserves language comprehension operates according to similar principles as memory in other domains.
© 2006 Elsevier Inc. All rights reserved.

Keywords: Language comprehension; Memory; Retrieval interference; Cue-overload; Parsing

Introduction

Sentence comprehension requires the ability to identify grammatical dependencies among words in a sentence that can often be separated from each other by a considerable distance, as in (1a) and (1b).

(1a) It was the client who the clerk at the reception desk forgot to announce to the manager.
(1b) It was the manager that the client who the clerk disliked was introduced to.

Intuition suggests that comprehenders must "hold on" to the noun phrase (NP) *the client* across intervening material until it can be integrated as the object of the verb *announce* in (1a) or *disliked* in (1b). As the resources available for sentence processing are often argued to be limited (e.g., Gibson, 1998; Just & Carpenter, 1992; Wanner & Maratsos, 1978), this intuition leads naturally to the idea that memory demands are an important determinant of difficulty in sentence comprehension. Indeed, researchers have proposed a variety of metrics for sentence complexity based on the hypothesized memory demands that sentences might impose upon comprehenders, including the number of intervening items (Gibson, 2000; Warner & Glass, 1987), embeddings (Miller & Chomsky, 1963; Miller & Isard, 1964), or incomplete dependencies (Abney & Johnson, 1991; Gibson, 1998; Kimball, 1973; Lewis, 1996).

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

0749-596X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2006.03.007

The assumption common to all these accounts is that comprehenders' ability to access a constituent (e.g., *the client*) at the point when it must be integrated with other constituents decreases as the amount of information retained in memory increases. Unfortunately, researchers have not fully articulated the memory mechanisms that comprehenders use to store and retrieve constituents during sentence processing, so it remains unclear why constituents might become less accessible as memory demands increase. However, basic research in memory retrieval offers several logical hypotheses. Cases such as (1a) subject the comprehenders' representation of the initial noun phrase to a classic form of *retroactive interference*, as the noun phrases must be maintained in memory while potentially interfering material is processed. The processing of the intervening material may reduce the encoding, storage, or retrieval of the noun phrase, or even displace the noun phrase from memory entirely. Alternatively, the representation of the noun phrase may simply decay over the time needed to process the interpolated material. Decay and displacement are not viable mechanisms in cases such as (1b), where the noun phrase *the manager* occurs before the "held" noun phrase *the client*. Structures in which other information must be retained in memory before the first constituent in a dependency is processed potentially subject the constituent to a form of *proactive interference*. In such cases, representations formed from early processing may adversely affect the encoding, storage, or retrieval of the constituent.

Although these hypotheses are all logically possible, investigations of capacity limits and forgetting in the domain of working memory have increasingly looked away from displacement or decay explanations and towards accounts that emphasize interference as the major constraint on accessing information in memory (e.g., Anderson & Neely, 1996; Crowder, 1976; see Nairne, 2002 for a recent review). Studies investigating the effects of semantic similarity on memory retention have provided particularly strong evidence implicating interference as the major determinant of forgetting. For example, recall or recognition of an item is reduced when it is semantically similar to other studied items, in circumstances where the reduction cannot be attributed to decay or displacement (e.g., Gorfein & Jacobson, 1972, 1973; Petrusic & Dillon, 1972; Watkins & Watkins, 1975; Wickens, 1970).

Additionally, research in this vein has suggested that interference has its primary effect on retrieval processes. Although similar items could interfere with the encoding or storage of a test item, the bulk of experimental evidence indicates that the locus of the interference effect is at retrieval, with little or no effect on memory encoding or storage (Dillon & Bittner, 1975; Gardiner, Craik, & Birstwistle, 1972; Tehan & Humphreys, 1996). For example, Gardiner et al. (1972) demonstrated the adverse effects of interference on retrieval using a "release from proactive interference (PI)" design (e.g., Watkins & Watkins, 1975; Wickens, 1970), in which PI is built up across trials by presenting subjects with lists from the same category and then "released" by switching semantic categories. Gardiner et al. presented participants with words from a subset of a category (e.g., garden flowers) during the build up trials and then switched to a complementary subset (e.g., wild flowers) on the release trial. All participants received the general category name (e.g., flower) as a cue at presentation in the first trial, but not in the following buildup trials. On the release trial, one experimental group received the subset cue (e.g., wild flowers) before the presentation, another experimental group received the cue at retrieval, and a control group did not receive any cue. Crucially, Gardiner et al. found that both the experimental groups showed similar amounts of release from PI. Because it was unlikely that the group that received the subcategory cue at retrieval changed their encoding of the release list, these findings suggest that PI negatively affected retrieval processes. Interference is thought to induce *cue-overload*, where the cues needed to recover an item become associated with many different items in memory and can not reliably elicit any single target (Anderson & Neely, 1996; Watkins & Watkins, 1975, 1976). In the study presented here, we investigated whether a similar principle is operative in processing sentences that are thought to impose high demands on memory resources.

Interference in sentence comprehension

The role of interference in sentence processing has only recently gained attention (Gordon et al., 2001, Gordon, Hendrick, & Johnson, 2004; Gordon, Hendrick, & Levine, 2002; Lewis, 1996; Van Dyke, 2002; Van Dyke & Lewis, 2003). In a series of studies investigating the role of various noun phrase types in sentences such as (2a) and (2b), Gordon and colleagues (Gordon et al., 2001, 2004) found that the classic processing advantage for subject-relative clauses over object-relative clauses was reduced or eliminated when the second noun phrase was either a pronoun (*you* or *everyone*) or a proper name (*Joe*). They attribute this reduction in processing cost to the reduction of similarity-based interference. Common noun phrases refer indirectly by virtue of their description, while pronouns and proper names refer directly, since their semantic value is provided extensionally with respect to specific objects in the discourse. Their data indicate that similarity-based interference occurs when the second noun phrase is from the same referential class as the first noun phrase, but it is reduced or eliminated when the noun phrases are from different classes.

(2a) The banker that praised [the barber \ a barber \ Joe \ you \ everyone] climbed the mountain.
(2b) The banker that [the barber \ a barber \ Joe \ you \ everyone] praised climbed the mountain.

In an important extension of this work, Gordon et al. (2002) investigated the role of interference with a memory load paradigm, which has the additional advantage of allowing for a more direct manipulation of the contents of memory during comprehension. In this study, subjects were presented with a memory load prior to reading the sentence, and asked to remember that list of words while reading. The size of the load was constant, but the similarity of items in the load vis-à-vis the linguistic material was manipulated. For example, memory load was either a list of three names or three role descriptions (e.g., banker), and the noun phrases in the sentence were either names or roles (different from those in the memory load). In this paradigm, potential for interference comes from matching or mismatching the load words to the words in the sentences (see 3a and 3b).

(3a) Matched memory load: Joel-Greg-Andy
Sentence: It was Tony that [liked Joey \ Joey liked] before the argument began.
(3b) Mismatched memory load: Joel-Greg-Andy
Sentence: It was the dancer that [liked the fireman \ the fireman liked] before the argument began.

Gordon et al. (2002) observed a clear match effect on accuracy to comprehension questions, in which the matched conditions produce significantly higher error rates than the mismatched conditions. This finding suggests that comprehension was difficult when there were salient noun phrases in memory that match the noun phrase types in the sentence. The effect was also seen in longer reading times in the matched conditions, suggesting that the presence of similar noun phrases has an online effect during syntactic processing.

While these studies serve to demonstrate the similarity-based interference effect, they do not provide clear insight into the underlying mechanism. Specifically, it is not possible to evaluate whether these effects occurred during the encoding of subsequent noun phrases or during retrieval of the clefted constituent. If a slowdown were observed at the point when the similar item is first encountered (i.e., in the region containing "the noun praised/praised the noun" in sentences (2a) and (2b) or "It was the noun phrase that" in sentences (3a) and (3b)), this would suggest an encoding effect, due to the presence of same noun phrase type distractors in memory. However, if a slowdown were observed in the region where the dependency is resolved (i.e., "climbed" in sentences (2a) and (2b) or "liked the fireman/the fireman liked" in (3a) and (3b)), it is more difficult to attribute the effect to encoding versus retrieval. Slowdowns in this region may arise if the quality of the memory representation is diminished or if the retrieval cues for a properly encoded representation are low in distinctiveness due to the presence of cue-overload.

The Gordon et al. (2001) data show what appears to be a similarity effect in the region when the similar item is first encountered, which might suggest an encoding explanation. However, the slowing of the reading time for a second noun phrase as compared with a pronoun or proper name in sentences such as (2a) and (2b) is confounded with both frequency and length effects. That this effect might simply reflect a speed up for pronoun or proper name conditions rather than a slowdown for the similarity conditions is supported by the fact that the similarity effect in the early region is significant for "the barber" vs. "you", where both frequency and length contribute to the reading time difference, but not significant for "the barber" vs. "Bob", where frequency may contribute less to the reading time difference (see, Gordon et al., 2001). In the memory load study, Gordon et al. (2002) also found similarity effects between "the barber" and "Bob" in pre-retrieval regions. However, this may be due to an encoding process encouraged by the dual-task paradigm, in which subjects specifically attempt to distinguish items in the sentence from items in the memory list. Matters are complicated further by the fact that Gordon et al. (2002) also found a similarity effect in the region of the dependency, which, as noted, could implicate either an encoding or a retrieval account.

Fedorenko, Gibson, and Rohde (2006) have recently extended the Gordon et al. (2002) findings, measuring the effects of low and high memory load (1-word versus 3-word lists) on reading times for subject- and object-relative clause constructions (e.g., *The physician who consulted the cardiologist checked the files in his office* versus *The physician who the cardiologist consulted checked the files in his office*). They report no effects of load at the first noun phrase, but reliably slower reading times for object-relative clauses under high load conditions. As with the Gordon et al. (2002) data, these findings indicate that similarity-based interference is an important determinant of comprehension difficulty, at least when the load is high so that interference is substantial. However, because the effects are found in the region that contains the dependency, they do not serve to uniquely motivate an encoding or retrieval account.

In order to cleanly discriminate between these accounts it is necessary to explicitly manipulate either the quality of encoding or retrieval cues available to the reader. A straightforward way to accomplish the latter is to employ retrieval cues that either do or do not uniquely identify the target noun phrase. This was not done in either the Gordon et al. studies or the Fedorenko et al. study, where the retrieval cues in all the manipulations were equally ambiguous across conditions. For example, in (2a) and (2b), "the barber" or "you" or "Bob" are just as plausible subjects for "climbed" as is the actual subject "the banker". Similarly, in (3a) and (3b), "the barber" and "Tony" make sense as either the subject or the object of the verb "liked." In the experiment reported here, we directly manipulated the verb so that nouns in the memory set were either plausible or implausible as a direct object of the verb. For example, in (1) in Table 1, none of the memory load items, *table*, *sink*, *or*, *truck*, are plausible as the object of the verb *sailed*. However, all are plausible as the objects of the verb *fixed* in example (2). If the interference observed in Gordon et al. (2002) study is due to an encoding effect, then the plausibility of the verb should not matter. However, if the interference has its locus at retrieval, then we expected that reading times around the verb in (1) would be slower than (2) because cues provided by the verb do not uniquely pick out the clefted noun phrase *the boat*. We included no-memory-load conditions such as (3) and (4) in Table 1 to insure that there were not other confounding differences between the two verbs.

Two recent studies suggest that manipulating the plausible competitors in memory should have measurable effects on reading times at the point at which the dependency is resolved. Van Dyke and Lewis (2003) compared sentences such as (4a) and (4b) in which a distracting noun phrase intervened between a long distance subject-verb dependency. In these sentences, the critical region for retrieval is the phrase *was complaining* in which the long distance subject *resident* must be retrieved despite a more recent noun phrase (*warehouse*). Increased reading times were observed in sentence (4b) as compared to sentence (4a), despite intervening regions of the same length. Van Dyke and Lewis suggested that the result was due to retrieval interference generated by the fact that the intervening noun phrase in (4b), which is a grammatical subject, matches the retrieval cues of the verb phrase (VP) *was complaining*, which is looking for its subject. In contrast, the intervening noun phrase in (4a) is the object of a preposition and hence does not match the subject cues of the verb.

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

(4b) The worker was surprised that the resident who said that the warehouse was dangerous was complaining about the investigation.

Further support for the retrieval interference theory comes from Van Dyke (2002; 2006), who crossed the syntactic manipulation in (4a) and (4b) with a plausibility manipulation to create two additional conditions, as in (4c) and (4d).

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

(4d) The worker was surprised that the resident who said that the neighbor was dangerous was complaining about the investigation.

This permitted a test of whether the semantic information from the verb was used during retrieval, and if so, whether the semantic fit between the critical verb (*was complaining*) and the intervening noun phrase *neighbor* was sufficient to cause interference, even when the noun phrase did not fit the grammatical cues from the verb (i.e., in (4c) *neighbor* is not a subject, but is the type of entity that can complain). In three experiments utilizing self-paced reading and eye-tracking measures, Van Dyke found slower reading times and lower comprehension accuracy when the intervening noun phrase was a plausible subject of the critical verb as compared to when it was not, even in the conditions when the intervening noun phrase did not match the grammatical cues of the verb. These results are not easily explained by an encoding account, since no differences at the manipulated noun phrase were observed, or by Gordon's account of interference as due to similarity versus mixing of noun phrases, since all the noun phrases were definite descriptions. Rather, the most natural explanation is that both the grammatical and semantic cues from the verb *was complained* are used during retrieval, and when the distracting noun phrases match these cues

Table 1
Sample experimental items

	Memory list	Sentence (interference manipulation underlined)	Comprehension question
(1) Non-interfering/Load	table-sink-truck	It was the boat/that the guy/who lived/by the sea/<u>sailed</u>/in two sunny days.	Did the guy live by the sea?
(2) Interfering/Load	table-sink-truck	It was the boat/that the guy/who lived/by the sea/<u>fixed</u>/in two sunny days.	Did the guy live by the sea?
(3) Non-interfering/No Load	—	It was the boat/that the guy/who lived/by the sea/<u>sailed</u>/in two sunny days.	Did the guy live by the sea?
(4) Interfering/No Load	—	It was the boat/that the guy/who lived/by the sea/<u>fixed</u>/in two sunny days.	Did the guy live by the sea?

they cannot be adequately discriminated from the desired target. If this is indeed the case, then we would expect that memory load in the current experiment would substantially slow processing when the retrieval cues provided by the verb match items in the memory set.

Method

Participants

Fifty-six students at New York University served as participants in the experiment. They were all native speakers of English and received $10 for their participation.

Design and materials

Four conditions comprised a 2 (Memory Load) $\times$ 2 (Interference) factorial design. The experimental sentences consisted of 36 object clefts, each with two versions differing only in the main verb (see Table 1). The manipulation of the main verb created interfering and non-interfering conditions vis-à-vis the words in the memory set, as these words were chosen so that they were either all plausible direct objects for the verb (interfering condition) or implausible direct objects for the verb (non-interfering condition). Plausibility was verified with a norming study involving 20 students from NYU who were not involved in the actual experiment and who were paid $10 for their time. These raters were asked to judge on a scale of (1 = not at all) to (7 = well) how well four words, the three memory words and the clefted word, fit into the simple main clause version of each experimental sentence. For example, the test items related to the experimental item in Table 1 would be as follows:

Non-interfering
The guy sailed the _____ in two sunny days.
Rated words: table, sink, truck, boat
Interfering
The guy fixed the ______ in two sunny days.
Rated words: table, sink, truck, boat

Results from this study showed that the three memory words were rated significantly lower in the non-interfering conditions than in the interfering conditions (1.62 versus 5.16, 95% CI = .26), $t(35) = -28.03$, $p < .001$. We also found that the word that would be the subject of the cleft (i.e., *boat* in the above example) in the experimental sentence was rated higher in the non-interfering condition than in the interfering condition (6.49 versus 5.89, 95% CI = .30), $t(35) = 4.10$, $p < .001$. In all cases, the subject of the cleft was rated higher than the memory words: (6.49 versus 1.62, 95% CI = .22), $t(35) = -44.06$, $p < .001$ for the non-interfering conditions, and (5.89 versus 5.16, 95% CI = .30) for the interfering conditions, $t(35) = -4.88$, $p < .001$.

In addition to the 36 experimental items, two types of filler items were included in the experiment so as to discourage participants from associating either the cleft construction or the presence of a memory list with the experimental conditions. Accordingly, one set consisted of 36 subject clefts, half of which also had an associated memory list, but all of these were unrelated to the verbs in the sentences they preceded (e.g., Memory List = *kettle*, *timeline*, *magnet*; Sentence = *It was the warden who worked in the state prison who discovered the escape tunnel in the prison.*) Comprehension questions for this set were divided evenly for "Yes" and "No" correct responses. The second set of filler items consisted of 108 non-clefted right-branching structures, half of which had associated memory lists which again, had no relationship to the verbs in the sentence (e.g., Memory List = *muffin*, *basin*, *theater*; Sentence = *The bartender commented that the patron grumbled that the drunkard enjoyed the booze.*) This set was also split evenly for "Yes" and "No" comprehension questions. These fillers together with the experimental items totaled 180 items that were shown to each participant.

Procedure

Four counterbalanced lists were created such that each experimental item occurred in only one condition in a list, and across lists every item occurred in all conditions. The lists were presented using the E-Prime experimental package (Schneider, Eschman, & Zuccolotto, 2002). In the Memory Load conditions, the subjects were first presented with a screen containing the three memory words, which was displayed for 3 s. The three words were presented simultaneously, centered on a single line and separated from each other by five "-" (dash) characters. Subjects were instructed to repeat the words out loud during this time and to remember them. Following this, they read the sentence using a self-paced phrase-by-phrase display. At the conclusion of the sentence, the comprehension question was displayed and students responded Yes or No using the "1" and "3" keys on the keyboard number pad. Following this, they were prompted to recall the memory words in order. Subjects typed each word into the computer using the keyboard, and were told to skip a position if they did not remember that word. Subjects were told to try to enter the words in the correct order, but if they could not, they could enter the words in any order. The next trial began immediately after entry of the last memory word (or entry of the enter key if they did not recall the word).

In the No Load conditions, subjects were first presented with a screen displayed for 3 s that said "No Memory Load." They then read the sentence using the same self-paced phrase-by-phrase display as in the Load conditions. The comprehension question appeared at the conclusion of the sentence, and the next trial began immediately following their response.

Results

The experiment produced three dependent measures: reading times during self-paced reading, accuracy to the comprehension question, and accuracy of recall for the memory words. Analyses of variance (ANOVA) were conducted on each measure using error terms based on participant (F_1) and item variability (F_2). The results of these ANOVAs are presented in Table 2, together with min-F′ statistics (Clark, 1973). For brevity, we report only those dependent measures where at least one F-statistic was greater than zero. For comparisons between condition means, we report 95% confidence intervals (CIs) based on the mean squared errors (MSE) of the associated effects from the participant analyses (reported in Table 2). These were calculated according to the procedure for within-participant confidence intervals described by Loftus and Masson (1994; Masson and Loftus, 2003). Planned contrasts were carried out using the 95% confidence intervals for pairwise differences between condition means.

Recall

Recall results were tallied according to both a strict criterion, which required the correct items in the correct order, and a lenient criterion, which permitted any order as long as the items were correct. There were only slight differences between these methods, and no significant effects in these data. According to the strict criterion, the proportion of correct responses in the non-interfering condition was .79 and .77 for the interfering condition, 95% CI = .04. According to the lenient criterion, the proportion of correct responses in the non-interfering condition was .80 and .78 for the interfering condition, 95% CI = .04.

Comprehension accuracy

Responses to the comprehension question revealed a significant effect of Memory Load, such that responses in the Load conditions were more accurate than in the No Load conditions (.83 versus .87 correct, 95% CI = .03). We also observed a significant effect of interference, such that responses in the non-interfering conditions were more accurate than in the interfering conditions (.87 versus .83, 95% CI = .03).

Reading times

Reading times were analyzed as six separate regions, demarcated by slashes in Table 1. The measure of most

Table 2
Analysis of variance results for all dependent measures with F > 1, with mean square error (MSE) for each effect

Measure	Main effect of load	Main effect of interference	Load × Interference
Comprehension accuracy	$F_1(1,55) = 6.14$, $p = .02$; $MSE_1 = .02$	$F_1(1,55) = 4.04$, $p = .05$; $MSE_1 = .02$	$F_1 < 1$
	$F_2(1,35) = 4.41$, $p = .04$; $MSE_2 = .01$	$F_2(1,35) = 8.41$, $p = .006$; $MSE_2 = .007$	$F_2 < 1$
	min $F'(1,78) = 2.56$, $p = .11$	min $F'(1,88) = 2.72$, $p = .10$	min $F' < 1$
Reading Time: Region 1	$F_1(1,55) = 5.23$, $p = .03$; $MSE_1 = 67914$	$F_1(1,55) = 1.95$, $p = .17$; $MSE_1 = 15390$	$F_1 < 1$
	$F_2(1,35) = 13.82$, $p = .001$; $MSE_2 = 16534$	$F_2(1,35) = 1.77$, $p = .19$; $MSE_2 = 10916$	$F_2 < 1$
	min $F'(1,85) = 3.79$, $p = .05$	min $F' < 1$	min $F' < 1$
Reading Time: Region 2	$F_1(1,55) = 1.23$, $p = .27$ $MSE_1 = 31574$	$F_1 < 1$	$F_1 < 1$
	$F_2(1,35) = 3.43$, $p = .07$ $MSE_2 = 7264$	$F_2 < 1$	$F_2 < 1$
	min $F' < 1$	min $F' < 1$	min $F' < 1$
Reading Time: Region 5 (Critical)	$F_1 < 1$	$F_1(1,55) = 4.34$, $p = .042$; $MSE_1 = 3800$	$F_1(1,55) = 4.07$, $p = .04$; $MSE_1 = 5790$
	$F_2 < 1$	$F_2(1,35) = 2.16$, $p = .15$; $MSE_2 = 4901$	$F_2(1,35) = 5.58$, $p = .02$; $MSE_2 = 2713$
	min $F' < 1$	min $F'(1,68) = 1.44$, $p = .23$	min $F'(1,90) = 2.35$, $p = .13$
Reading Time: Region 6	$F_1(1,55) = 15.81$, $p < .001$; $MSE_1 = 49787$	$F_1 < 1$	$F_1 < 1$
	$F_2(1,35) = 18.90$, $p < .001$; $MSE_2 = 26762$	$F_2 < 1$	$F_2 < 1$
	min $F'(1,88) = 8.61$, $p < .01$	min $F' < 1$	min $F' < 1$

interest is the reading time in the region of the manipulated verb (Region 5), as all other regions were identical across conditions. Reading times were trimmed to 2.5 standard deviations of the mean, a procedure that affected 2.5% of the data.

Table 3 presents the reading times for all six regions. In Region 1, the effect of Load was significant, with the Load condition being significantly slower than the No Load condition (897 ms versus 818 ms, 95% CI = 69). No other effect was significant in this region, $Fs \approx 1$. Likewise, in Regions 2–4 no effects were significant, $Fs \approx 1$, except for a trend for an effect of Load in the items analysis in Region 2 with the Load condition being significantly slower than the No Load condition (786 ms versus 760 ms, 95% CI = 28).

In Region 5, the critical region containing the interference manipulation, the effect of Load was not significant, $Fs < 1$. The effect of Interference was significant by subjects such that the non-interfering conditions were faster than the interfering conditions (554 ms versus 571 ms, 95% CI = 16). This effect did not reach significance by items. The interaction of Load × Interference was significant, both by subjects and by items. Planned comparisons reveal that this is due primarily to a significant interference effect in the Load conditions, with an effect size of 38 ms, 95% CI = 27. The effect of interference was not present in the No Load conditions, effect size of −3 ms, 95% CI = 25. Planned comparisons also suggest a marginal effect of Load in the non-interfering conditions, with an effect of 27 ms, 95% CI = 26.7. This was due to the Load condition being read more quickly (540 ms) than the No Load condition (567 ms). No other pairwise effects were significant, $Fs \approx 1$.

In Region 6, the final region, we observed a main effect of Load, such that the Load conditions were read significantly faster than the No Load conditions (1257 ms versus 1375 ms, 95% CI = 59). No other effects were significant in this region, $Fs < 1$.

General discussion

Summary

The comprehension accuracy results confirmed that the Load condition was more difficult than the No Load condition, as subjects were less accurate when there was a memory load. This is consistent with the reading time results, which show significant effects of Load in regions 1 and 6, and a trend for an effect in Region 2. The numerical pattern of these effects suggests that subjects read more slowly at the beginning of the sentence (compare Regions 1 and 2 where load conditions were slower), possibly reflecting attempts to rehearse the memory items. From Region 3 onward, except for critical Region 5, the load conditions were numerically faster, suggesting that subjects were speeding up in order to quickly get to the recall task. This interpretation is consistent with the lower comprehension accuracy in these conditions, which would be expected if the subjects paid less attention to the reading task under dual-task conditions.

The significant effect of interference in the comprehension results accords with the results of our plausibility norming: the clefted noun was rated as a more plausible object of the manipulated verb in the non-interfering conditions compared with the interfering conditions. This higher plausibility may have led to greater ease at understanding the clefted version of the non-interfering sentence, which produced higher comprehension accuracy as compared with the interfering conditions.

Reading times for Region 5, the locus of the interference manipulation, provided the critical test of our retrieval interference hypothesis. We found clear support for retrieval interference from the significant effect of interference and the significant interaction in this region, which revealed that the interference effect was linked to the difference between the two sentence types in the Load conditions. Pairwise comparisons showed no difference between the interference conditions when there was no load present. However, reading was slowed in this region when the distracting referents in memory matched the retrieval cues inherent in the verb seeking its object (Interfering/Load condition), relative to conditions in which the distracting referents did not match the verb's retrieval cues (Non-interfering/Load condition) or when there were no distracting referents at all (No Load conditions). This result is not fully consistent with the account of Gordon et al. (2002), who suggested that the source of interference was the presence in memory of distracting items of the *same noun phrase type* as those in the sentence. As our distracting items were always of the

Table 3
Reading times (ms), standard error in parenthesis

Condition	Region					
	1	2	3	4	5 (Critical)	6
Non-interfering/Load	908 (40)	788 (38)	655 (22)	678 (25)	540 (15)	1264 (55)
Interfering/Load	887 (44)	785 (36)	664 (24)	688 (24)	578 (20)	1249 (52)
Non-interfering/No Load	830 (34)	766 (26)	674 (20)	692 (22)	567 (17)	1369 (50)
Interfering/No Load	805 (31)	754 (24)	672 (20)	687 (21)	564 (16)	1381 (50)

same noun phrase type as those in the sentence (i.e., descriptive noun phrases), our results indicate that interference was defined with respect to the retrieval cues of the verb, and not the semantic properties of the items stored in memory. However, as we did not explicitly manipulate the similarity among the encoded items, only the similarity of the encoded items to the verb, we cannot completely rule out the possibility that semantic similarity might have some adverse effects on the encoding of the critical items in memory.

Retrieval interference in sentence processing

The current experiment supports a retrieval-based account of interference effects in sentence processing, one that is compatible with the hypothesis that a cue-based retrieval mechanism mediates the creation of grammatical dependencies during parsing. One such mechanism has been proposed in Van Dyke and Lewis (2003; see also Lewis and Vasishth, 2005; Van Dyke, 2002), in which parsing success depends on the extent to which required constituents can be retrieved from working memory. On this account grammatical heads provide retrieval cues that are used to access previously stored items via a content-addressable retrieval process (McElree, 2000, 2006; McElree, Foraker, & Dyer, 2003).

The defining property of a content-addressable retrieval mechanism is that information (cues) in the retrieval context enables direct access to relevant memory representations, without the need to search through extraneous representations. This type of mechanism can be implemented in models with diverse storage architectures, including those with localized representations and those with highly distributed representations (Clark & Gronlund, 1996). Importantly, empirical measures of the dynamics of retrieval suggest that content addressability is a very general property of human memory, and that this type of process mediates the recovery of items from both short- and long-term memories (for a review, see McElree, 2006).

Following several models proposed in the memory literature (e.g., Gillund & Shiffrin, 1984; Hintzman, 1984, 1988; Nairne, 1990), we assume that cues in the retrieval context are combined multiplicatively to produce a single retrieval probe. For example, in the Search of Associative Memory (SAM) model of Gillund and Shiffrin (1984), all cues are combined into a retrieval probe, which gives the strength of the relationship between each possible probe cue ($Q_1, \ldots, Q_m$) and the memory trace, assumed to be a vector of features that can differ in value and type, for each item stored in memory. The probability, P, of retrieving an item, I_i, is a function of the strength of association S between each probe cue, Q_j, and the features of the actual memory trace, denoted as $S(Q_j, I_i)^{w_j}$ where w_j is a weighting factor denoting the relative saliency of the different cues. Specifically, this probability is defined in the equation below:

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

Thus, the probability of retrieving a particular item given this retrieval probe is an increasing function of the degree of the match between the probe and the memory item, or the probe-to-item strength, shown by the numerator. Crucially, however, because retrieval is mediated by the resonance between the cues and items in memory, the probability of retrieving a particular item will be a *decreasing* function of the degree to which the probe matches *other* items in memory, or the sum of the probe-to-item strengths for all other items, shown by the denominator. Thus, when a competitor in memory is very similar to the target item, there is an increased probability that it will be incorrectly retrieved (viz., its probe-to-item strength will be very similar to the probe-to-target strength). Therein lies the defining condition for similarity-based interference, which we assume was operative in our Interfering/Load condition.

In accordance with the framework outlined in Van Dyke and Lewis (2003; see also Lewis and Vasishth, 2005; Van Dyke, 2002), we assume that the retrieval cues used to recover a representation of the relevant clefted item in our experimental sentences were largely derived from both the linguistic properties of the head verb, including subcategorization and thematic role information, and experience-based information associated with the verb's lexical entry that reflects the plausibility of a particular grammatical association. Lewis and Vasishth (2005) present an implemented computational model of cue-based retrieval that follows Van Dyke and Lewis (2003) in utilizing primarily syntactic cues during retrieval. However, Van Dyke (2002; 2006) and the experiment reported here suggest that information associated with the meaning of competing items is also used during retrieval. Fully specifying all the cues that might drive memory retrieval is beyond the scope of our current understanding, as it requires enumerating and investigating all forms of information that might be computed during sentence processing. However, general claims about cue-based retrieval in parsing do not depend on any particular assumed set of features. Rather, the relevance of a cue-based approach is motivated by demonstrations that comprehenders have available some set of features that capture the distinctions necessary for identifying various grammatical dependencies within the sentence (Lewis & Vasishth, 2005; Van Dyke, 2002; Van Dyke & Lewis, 2003) and by demonstrations that comprehension is sensitive to factors that are hypothesized to be intrinsic to the cue-based operation, such as the similarity-based interference documented here.

An important consequence of the fact that retrieval is determined not only by the similarity of the memory probe to the target item, but also by the similarity of the probe to all other items in memory is that the memory probe created from the available retrieval cues must be able to *distinguish* the target from other matching items in memory (Crowder, 1976; Nairne, 2001). We suggest that it is the relative ability to distinguish among alternatives that explains the Gordon et al. (2001, 2004) results, even though those studies were not designed to explicitly manipulate retrieval cues. The Gordon et al. (2001, 2004) findings are important to consider in detail because they could be interpreted as implicating encoding or storage rather than retrieval for the adverse effects of interference.

As noted above, Gordon et al. (2001, 2004) found that interference was reduced when the second noun phrase in a relative clause (examples 2a and 2b above) was either a pronoun (*you*) or a proper name (*Joe*) as compared to a noun phrase (*the barber*). We noted that interference is present at retrieval with all three types of noun phrases because all are just as plausible subjects for the embedded verb (*climbed*) as is the actual subject (*the banker*). This might lead one to assume that a retrieval account predicts no difference between these conditions. Crucially, however, Gordon et al. (2004) point out that these noun phrase types have different referential properties. It is this difference that we presume differentiates the three types of structures at retrieval. Although the retrieval cues at the verb remain the same, and so the match of the probe to the target (*the banker*) remains constant, the similarity of the probe to other items in memory would not be expected to remain constant across the different types of second noun phrases. Hence, the Gordon et al. (2001, 2004) findings can be explained with a simple assumption that the match of the probe to a pronoun or to proper name are lower than to a common noun phrase, perhaps because semantic properties associated with the latter are inherently richer and therefore more distinctive than properties associated with the former. The net effect of this difference among the competitors in memory will be to increase the probability of recovering the target noun phrase when paired with either a pronoun or proper noun. In terms of the equation cited above, a lower match of the probe to a pronoun or proper noun will decrease the value of the denominator, which will result in a higher probability of retrieval, even though the numerator, reflecting the match of the probe to the target NP, remains constant.

General implications

There has been considerable debate about whether the language system has a dedicated working memory system, or whether it utilizes the same system as other non-language skills (e.g., Caplan & Waters, 1999). The evidence presented here, and in other investigations of similarity-based interference in sentence processing, indicates that interference plays a defining role in determining whether comprehension will be successful. Importantly, our data localize these interference effects to retrieval, just as basic memory research has demonstrated that both proactive and retroactive interference adversely affect retrieval (e.g., Anderson & Neely, 1996; Crowder, 1976; Nairne, 2002). These findings suggest that, in terms of function, the memory system that subserves language comprehension operates according to similar principles as memory in other domains. At this level of analysis, there is no compelling evidence to suggest that language comprehension is not completely circumscribed by the memory mechanisms at work in various non-language tasks.

Acknowledgments

This research was supported by National Science Foundation grant (BCS-0236732) awarded to BM and by a National Research Service Award (NIH F32-HD049215-01) to JV. We thank Serafina Shishkova for her assistance in generating and norming the material used in the experiment.

References

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

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork & R. A. Bjork (Eds.). *Handbook of perception and memory, Memory* (Vol. 10, pp. 237–313). San Diego: Academic Press.

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

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.

Dillon, R. F., & Bittner, L. A. (1975). Analysis of retrieval cues and release from proactive inhibition. *Journal of Verbal Learning and Verbal Behavior, 14*, 616–622.

Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language, 54*(4), 541–553.

Gardiner, J. M., Craik, F. I. M., & Birstwistle, J. (1972). Retrieval cues and release from proactive inhibition. *Journal of Verbal Learning and Verbal Behavior, 11*, 778–783.

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, 2*(6), 1411–1423.

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

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). Memory search in a Brown-Peterson short-term memory paradigm. *Journal of Experimental Psychology, 99*, 82–87.

Gorfein, D. S., & Jacobson, D. E. (1973). Proactive effects in short-term recognition memory. *Journal of Experimental Psychology, 95*, 211–214.

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

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 9*(4), 528–551.

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

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

Lewis, R. L. (1996). Interference in short-term memory: the magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research, 2*(1), 93–115.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 2*(3), 375–419.

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

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

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

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language, 4*(1), 67–91.

Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology, 5*(3), 203–220.

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

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

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

Nairne, J. S. (2001). A functional analysis of primary memory. In I. H. L. Roediger, J. S. Nairne, I. Neath, & A. M. Suprenant (Eds.), *The nature of remembering: essays in honor of Robert G. Crowder*. Washington, DC: APA.

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory, 1*(5/6), 389–395.

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

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime User's Guide. Psychology Software Tools, Pittsburgh, PA.

Tehan, G., & Humphreys, M. S. (1996). Cuing effects in short-term recall. *Memory and Cognition, 24*, 719–732.

Van Dyke, J.A. (2006). Interference effects from grammatically unavailable constituents during sentence processing, unpublished manuscript.

Van Dyke, J.A. (2002). Retrieval Effects in Sentence Parsing and Interpretation. Unpublished Ph.D. Dissertation, University of Pittsburgh, Pittsburgh.

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, 4*(3), 285–316.

Wanner, E., & Maratsos, M. (1978). An ATN approach in comprehension. In M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 119–161). Cambridge, MA: MIT Press.

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

Watkins, M. J., & Watkins, O. C. (1976). Cue-overload theory and the method of interpolated attributes. *Bulletin of the Psychonomic Society, 7*(3), 289–291.

Watkins, O. C., & Watkins, M. J. (1975). Build-up 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.