Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Memory and Language 54 (2006) 541–553

Journal of Memory and Language

www.elsevier.com/locate/jml

The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources ☆

Evelina Fedorenko *, Edward Gibson, Douglas Rohde

Massachusetts Institute of Technology, USA

Received 8 April 2005; revision received 4 December 2005
Available online 15 February 2006

Abstract

This paper reports the results of a dual-task experiment which investigates the nature of working memory resources used in sentence comprehension. Participants read sentences of varying syntactic complexity (containing subject- and object-extracted relative clauses) while remembering one or three nouns (similar to or dissimilar from the sentence-nouns). A significant on-line interaction was found between syntactic complexity and similarity between the memory-nouns and the sentence-nouns in the three memory-nouns conditions, such that the similarity between the memory-nouns and the sentence-nouns affected the more complex object-extracted relative clauses to a greater extent than the less complex subject-extracted relative clauses. These results extend Gordon, Hendrick, and Levine's (2002) report of a trend of such an interaction. The results argue against the domain-specific view of working memory resources in sentence comprehension (Caplan & Waters, 1999).

Keywords: Working memory; Sentence comprehension; Domain specificity

Introduction

A major question in cognitive science concerns the nature and the functional organization of the working memory system. In psycholinguistic research, this question has focused on investigating the nature of the working memory resources underlying language processing. More generally, the question of the functional organization of the working memory system is relevant to the modularity debate (Fodor, 1983), which is aimed at understanding whether there exist cognitive modules—subserved by highly specialized neural structures—dedicated to specific cognitive functions [e.g., linguistic knowledge representation (e.g., Chomsky, 1986), face perception (e.g., Kanwisher, McDermott, & Chun, 1997), musical processing (e.g., Peretz & Hyde, 2003; McDermott & Hauser, in press)], or whether our cognitive system is more domain-general in nature, such that the same neural/cognitive resources are used for multiple cognitive functions.

Earlier research has suggested that different pools of working memory resources are used for processing visuo-spatial information and verbal information (e.g., Baddeley & Hitch, 1974; Baddeley, 1986; Hanley,

☆ This research was supported by NIH Grant 5R01DC05237-05 to the second author and NIH NRSA Grant 1-F32-MH65105-02 to the third author.
* Corresponding author.
E-mail address: evelina9@mit.edu (E. Fedorenko).

0749-596X/$ - see front matter
doi:10.1016/j.jml.2005.12.006

Young, & Pearson, 1991; Jonides et al., 1993; Shah & Miyake, 1996; Vallar & Shallice, 1990. Caplan & Waters (1999) have hypothesized that the verbal working memory pool can be further divided into two sub-pools: (1) verbal working memory for linguistic processing; and (2) verbal working memory for non-linguistic verbally-mediated cognitive tasks. In contrast, other researchers (e.g., Just & Carpenter, 1992; King & Just, 1991) have argued that linguistic processing and non-linguistic verbally-mediated cognitive tasks rely on the same pool of verbal working memory resources. This paper attempts to empirically evaluate these alternatives.

Two approaches have been traditionally used to address the question of working memory resources used in on-line linguistic processing: (1) an individual-differences approach, and (2) a dual-task approach. In the individual-differences approach, participants are divided into two or more groups on the basis of their performance on some form of a verbal working memory task and tested on linguistic structures of varying syntactic complexity. In the dual-task approach, on the other hand, participants perform two tasks simultaneously: (1) on-line sentence processing, and (2) a non-linguistic verbally-mediated task. The underlying assumption of the two approaches is that syntactic complexity should interact with group-type or with the difficulty of the secondary task, respectively, only if the non-linguistic verbally-mediated task and on-line linguistic processing rely on the same pool/overlapping pools of verbal working memory resources.

King and Just (1991) and Just and Carpenter (1992) provided suggestive evidence in support of the same resource pool/overlapping resource pools hypothesis.[1] This evidence consisted of differential behavior of low- and high-span readers, classified using Daneman and Carpenter's (1980) reading span task, in the processing of syntactic structures of low and high complexity (subject- vs. object-extracted relative clauses). However, Caplan and Waters (1999) noted that the required statistical analyses—interactions between group-type, syntactic complexity, and sentence region—were not reported, and the qualitative pattern of the reported data did not support the overlapping resource pools hypothesis. Furthermore, Caplan and Waters attempted to replicate King and Just's and Just and Carpenter's results using a variety of methods and large subject pools, and were not able to demonstrate the required interactions, nor were there any suggestions of such effects.

Waters and colleagues (Waters, Caplan, & Hildebrandt, 1987, 1995) also tested the overlapping resource pools hypothesis by conducting a series of experiments using a dual-task approach where subjects were asked to perform self-paced reading/listening while maintaining a memory load (usually, a string of digits). No on-line interactions or suggestive trends between syntactic complexity and memory load were found in any of the experiments. Waters et al. interpreted their results as supporting the hypothesis whereby there is an independent pool of verbal working memory resources dedicated to on-line sentence processing (for a more complete review of the studies outlined above, see Caplan & Waters, 1999). In addition to the individual-differences studies and the dual-task experiments, Caplan and Waters (1999) reported some data from neuropsychological studies conduced with various patient populations. These data are interpreted as providing further support for the idea of an independent pool of verbal working memory resources for on-line linguistic processing (see Caplan & Waters, 1999, pp. 87–92).

It is worth noting that there have been several reports of off-line interactions between syntactic complexity and memory load in the literature. For example, Waters et al. (1987) and Waters and Caplan (1996) found that syntactic complexity had an effect on the number of sentence-final words recalled in a sentence-acceptability-judgment task. Similarly, Wanner and Maratsos (1978) used a task where sentence presentation was interrupted by a list of words, which had to be recalled at the end of the sentence. They reported poorer word recall performance in more complex object-extracted relative clauses, compared to less complex subject-extracted relative clauses. Caplan and Waters (1999) used two different lines of argumentation to show that the off-line interactions observed in some of the previous experiments are still consistent with the idea of an independent pool of verbal working memory resources dedicated to on-line sentence comprehension. First, they made a distinction between interpretive (on-line) and post-interpretive (off-line) processes, which are involved in sentence comprehension. Interpretive processing, according to Caplan and Waters, involves the "extraction of meaning from a linguistic signal" (p. 79), whereas post-interpretive processing involves using this extracted meaning to accomplish tasks, like reasoning, planning actions, and storing information in long-term semantic memory. Caplan and Waters then argued that the off-line interactions observed between linguistic processing and non-linguistic verbally-mediated tasks do not directly address the question of an overlap in verbal working memory resources, because post-interpretive processing (used in off-line tasks) involves a variety of cognitive processes beyond linguistic processing. Second, Caplan and

[1] In fact, Just and Carpenter (1992) argued for a capacity-constrained comprehension model, where on-line language processing and non-linguistic verbally-mediated tasks rely on the *same* pool of resources. However, using the individual-differences approach and the dual-task approach, it is logically impossible to determine from the observed interactions the extent of the overlap—partial vs. complete—between the verbal working memory pools used for linguistic processing and other non-linguistic verbally-mediated tasks.

Waters argued that because the only experiments in which interactions between syntactic complexity and memory load have been observed involved one task interrupting the other task, it is likely that such interactions resulted from the necessity to shift attention back and forth between the two tasks, rather than from an overlap in verbal working memory resource pools between the two tasks.

More recently, Gordon et al. (2002) argued that the load manipulation used in the previous dual-task experiments (e.g., increasing the number of memory-items in the digit-span task) was not the right one for the purposes of assessing the nature of verbal working memory resources in sentence comprehension. They suggested that the critical characteristic of the memory load is its representational nature in relation to the representational nature of the linguistic materials. Specifically, Gordon et al. argued that working memory capacity in language processing should be conceptualized not in terms of *the number of items* that must be kept active in memory during the comprehension process, as has been suggested by Daneman and Carpenter (1980), King and Just (1991), Gibson (1998, 2000), and Lewis (1996) among others, but rather in terms of *the amount of interference* produced by the items that must be kept active in memory.

Gordon et al. tested the overlapping resource pools hypothesis of verbal working memory for sentence comprehension using a novel dual-task paradigm, where participants read sentences of high and low syntactic complexity (subject- and object-extracted cleft sentences), which contained either occupations (e.g., "It was the dancer that liked the fireman/that the fireman liked before the argument began"), or personal names (e.g., "It was Tony that liked Joey/that Joey liked before the argument began"). At the same time, participants were asked to remember a set of three words, which could also be either occupations (e.g., poet, cartoonist, voter), or personal names (e.g., Joel, Greg, Andy). This design resulted in two match conditions (memory-nouns and sentence-nouns from the same category) and two non-match conditions (memory-nouns and sentence-nouns from different categories). At the end of each sentence, participants were asked to answer a comprehension question about the content of the sentence and to recall the words from the memory task. Gordon et al. hypothesized that the similarity between the memory-nouns and the sentence-nouns might affect the more complex sentences (object-extracted clefts) to a larger extent.

To explain their results, Gordon et al. adopted the similarity-based interference framework (e.g., Gillund & Shiffrin, 1984; Hintzman, 1986). In this framework, interference effects in sentence comprehension are argued to apply at the retrieval stage of the memory process and are conceptualized in terms of an overlap in retrieval cues. Specifically, it is argued that with an increase in the overlap in retrieval cues for different items in memory, the cue-to-target strength for any individual item decreases, making the retrieval process more costly. With regard to the 2002 experiment, Gordon et al. argued that in cases where the memory traces of the memory-nouns are similar to the memory trace of the relevant antecedent, interference takes place, such that it is harder to identify the relevant antecedent among all the available memory traces. Gordon et al. further hypothesized that these effects might be larger in object-extracted clefts due to a higher memory demand posed by these structures, compared to the subject-extracted clefts. The most interesting result of Gordon et al.'s experiment is a reliable interaction between syntactic complexity and noun type match in comprehension question accuracy data, such that there was a larger difference between subject- and object-extracted clefts for the match conditions than for the non-match conditions. The authors interpreted these results as evidence against an independent verbal working memory resource pool for sentence comprehension.

Gordon et al.'s (2002) results are the first report of an interaction between syntactic complexity and memory load in a paradigm where the two tasks did not interrupt each other. Thus, these results are not likely to be attributable to shared attention costs associated with task-switching (as has been proposed by Caplan & Waters, 1999). However, Gordon et al.'s results do not speak directly to the nature of verbal working memory resources in on-line (interpretive) linguistic processing, because the only significant interaction that Gordon et al. observed was an effect in response accuracies to comprehension questions. Although the on-line reading time data showed a trend towards a similar interaction, it was short of significance ($p = 0.13$ in the subjects analysis; $p = 0.19$ in the items analysis) and thus difficult to interpret. As described above, Caplan and Waters (1999) have argued that off-line tasks, like answering comprehension questions, tap into post-interpretive processing rather than interpretive processing. Therefore the question of the nature of working memory resources for on-line sentence comprehension is not yet resolved. The goal of this paper is to provide online evidence relevant to the question of domain specificity of working memory resources involved in sentence comprehension.

Experiment

This experiment was similar in design to Gordon et al.'s (2002) study. Participants read sentences phrase-by-phrase, and at the same time were required to remember one or three nouns, which were either similar to or dissimilar from the nouns used in the sentences. The design was different from that of Gordon et al.'s in several respects. First, we chose to use structures with subject- and object-extracted relative clauses, as opposed to clefts. Second, we only varied the noun type of the memory-nouns, keeping the nouns in the sentences the same. Third, we included a load manipulation in terms of the number of memory-nouns (one memory-noun vs. three memory-nouns). As discussed above, Gordon et al. (2002) proposed that working memory capacity in language processing should be conceptualized in terms of the amount of interference produced by the items that must be kept active in memory. Gordon et al. manipulated the amount of interference produced by the memory items by varying the degree of similarity between the memory-nouns and the sentence-nouns. It is plausible that the amount of interference is a function of not only the representational characteristics of memory items but also the number of memory items. Syntactic complexity may therefore interact with memory load (manipulated in terms of the number of items) in the context of similar elements. If this is the case, then we would expect a three-way interaction among the three factors, such that an interaction between syntactic complexity and the number of memory-nouns should be observed in the conditions where the memory-nouns and the sentence-nouns are similar, but not in the conditions where they are dissimilar.

Furthermore, there was a difference in the procedure, such that unlike Gordon et al. who used center-screen presentation, we used a moving-window presentation, which is arguably more natural, with a stronger resemblance to normal reading (Just, Carpenter, & Woolley, 1982). We reasoned that it was possible that part of the reason that Gordon et al. did not get an online interaction might be due to the procedure they used, which might not be sensitive enough.

Methods

Participants

Forty-four participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purposes of the study.

Design and materials

The experiment had a 2 × 2 × 2 design, crossing syntactic complexity (subject-extracted relative clause, object-extracted relative clause), memory load (one noun, three nouns), and memory-noun type (occupation, name). The nouns in the sentences were always occupations, and thus the memory-noun(s) could either match or not match the sentence-nouns in type.

The language materials consisted of 32 sets of sentences, having four different versions as in (1):

(1) a. *Subject-extracted*, *version 1*:
The physician | who consulted the cardiologist | checked the files | in the office.
b. *Subject-extracted*, *version 2*:
The cardiologist | who consulted the physician | checked the files | in the office.
c. *Object-extracted*, *version 1*:
The physician | who the cardiologist consulted | checked the files | in the office.
d. *Object-extracted*, *version 2*:
The cardiologist | who the physician consulted | checked the files | in the office.

As described above, there were two levels of syntactic complexity—subject- and object-extractions—but there were four versions of each sentence to control for potential plausibility differences between the subject- and object-extracted versions of each sentence. As a result, no independent plausibility control is needed in this design. Each participant saw only one version of each sentence, following a Latin-Square design (see Appendix A for a complete list of linguistic materials). The nouns for the memory task—the occupations and the names—were matched for frequency (using a Usenet corpus of 1.2 billion words) and length in syllables (the means are presented in Appendix B) and paired with the sentences, such that the memory-nouns were not related semantically (for occupations) or phonologically to the nouns in the sentence.

In addition to the target sentences, 40 filler sentences with various syntactic structures other than relative clauses were included. The filler sentences were preceded by one, two or three memory-noun(s), which were a combination of occupations and names. The length and syntactic complexity of the filler sentences was similar to that of the target sentences. The stimuli were pseudo-randomized separately for each participant, with at least one filler separating the target sentences.

Procedure

The task was self-paced phrase-by-phrase reading with a moving-window display (Just et al., 1982). The experiment was run using the Linger 2.85 software by Doug Rohde (available at http://tedlab.mit.edu/~dr/Linger/). Each experimental sentence had four regions (as shown in (1a–d)): (1) a noun phrase, (2) a relative clause (subject-/object-extracted), (3) a main verb with a direct object (always an inanimate noun phrase), and (4) an adjunct prepositional phrase. The memory-noun(s) was/were presented in capital letters in the center of the screen. Each trial began with the memory-noun(s) appearing on the screen for 600 ms (one noun) or 1800 ms (three nouns). Participants were instructed to try to remember the noun(s) as well as they could. This was followed by a blank screen for 500 ms, which in turn was followed by a series of dashes marking the length and position of the words in the sentence. Participants pressed the spacebar to reveal each region of the sentence. As each new region appeared, the preceding region disappeared. The amount of time the participant spent reading each region was recorded as the time between key-presses. To make sure the participants performed the memory task, a box appeared on the screen after the last region of the sentence was read, and the participants were instructed to type in as many of the nouns that were presented at the beginning of the trial as possible in any order. If the noun(s) were typed in correctly, the word "RIGHT" flashed briefly on the screen. If two of the three nouns were typed in correctly (in the hard-load conditions), the words "ALMOST RIGHT" flashed briefly. Finally, if the noun was typed in incorrectly (in the easy-load conditions) or if less than two nouns were typed in correctly (in the hard-load conditions), the word "WRONG" flashed briefly on the screen.

To make sure the participants read the sentences for meaning, two true-or-false statements were presented sequentially after the memory task, asking about the propositional content of the sentence they just read. Participants pressed one of two keys to respond "true" or "false" to the statements. After a correct answer, the word "CORRECT" flashed briefly on the screen, and after an incorrect answer, the word "INCORRECT" flashed briefly.

Participants were instructed to read sentences silently at a natural pace and to be sure that they understood what they read. They were also told to take wrong answers as an indication to read more carefully.

Before the experiment started, a short list of practice items and questions was presented in order to familiarize the participants with the task. Participants took approximately 35 min to complete the experiment.

Results

Memory task

The performance on the memory task was calculated using the following formula: hits/(hits + misses + false-alarms). This formula allowed us to give partial credit for partial responses in the hard-load conditions and to penalize participants for guessing. Minor spelling mistakes (deviations from the targets by up to two letters) were not taken into consideration and the words with such mistakes were counted as hits. Across conditions, participants performed at 86.7% correct. Table 1 presents the mean performance across the eight conditions of the experiment. A three-factor ANOVA crossing syntactic complexity (subject-extracted relative clause, object-extracted relative clause), memory load (one memory-noun, three memory-nouns), and memory-noun type (match, non-match) was performed. The analysis revealed a main effect of memory load ($F1(1,43) = 28.1$; MSe = .618; $p < .001$; $F2(1,31) = 28.6$; MSe = .453; $p < .001$; $\min F'(1,72) = 14.1$; $p < .001$), such that people had higher accuracy rates in the easy-load conditions (90.9%), compared to the hard-load conditions (82.5%). There was also a main effect of memory-noun type ($F1(1,43) = 16.8$; MSe = .381; $p < .001$; $F2(1,31) = 17.8$; MSe = .274; $p < .001$; $\min F'(1,73) = 8.64$; $p < .005$), such that names were recalled with higher accuracy (90.01%) than occupations (83.3%). No other significant effects were observed (Fs < 1.5).

Comprehension question performance

There were two comprehension questions following each experimental trial. Participants answered the first question correctly 79.4% of the time, and the second question 79.9% of the time. The percentages of correct answers by condition were very similar for the two questions, so we collapsed the results in our analyses. Table 2 presents the mean accuracies across the eight conditions of the experiment. A three-factor ANOVA crossing syntactic complexity (subject-extracted relative clause, object-extracted relative clause), memory load (one memory-noun, three memory-nouns), and memory-noun type (match, non-match) on the responses to the comprehension questions revealed a main effect of memory load ($F1(1,43) = 42.02$; MSe = 1.12; $p < .001$; $F2(1,31) = 63.4$; MSe = .794; $p < .001$; $\min F'(1,74) = 25.2$; $p < .001$), such that people answered comprehension questions more accurately in the easy-load conditions (85.3%), compared to the hard-load conditions (74.04%), but no other significant effects. All Fs were less

Table 1
Memory task performance in percent correct, as a function of syntactic complexity, memory load, and memory-noun type (standard errors in parentheses)

Memory load	Syntactic complexity			
	Subject-extraction (easy)		Object-extraction (hard)	
Memory-noun	Match	Non-match	Match	Non-match
One noun (easy)	86.7 (3.0)	93.8 (1.7)	88.1 (3.0)	95.2 (1.7)
Three nouns (hard)	80.8 (2.9)	85.8 (2.5)	78.2 (2.9)	85.3 (2.4)

Table 2
Comprehension accuracies in percent correct, as a function of syntactic complexity, memory load, and memory-noun type (standard errors in parentheses)

Memory load	Syntactic complexity			
	Subject-extraction (easy)		Object-extraction (hard)	
Memory-noun	Match	Non-match	Match	Non-match
One noun (easy)	83.2 (3.2)	86.7 (2.7)	86.9 (2.3)	84.4 (3.0)
Three nouns (hard)	77.6 (3.0)	73.6 (3.0)	70.3 (3.8)	74.7 (3.5)

than 1.5, except for a marginal three-way interaction ($F1(1,43) = 3.38$; MSe = .113; $p = .073$; $F2(1,31) = 2.87$; MSe = .0799; $p = .10$; $\min F'(1,70) = 1.55$; p =.22). The trend for this three-way interaction is likely due to the fact that numerically, the effect of memory load affected the object-extracted relative clause conditions more, and this trend appeared more pronounced in the match conditions. It is worth noting that we did not replicate Gordon et al.'s (2002) off-line interaction between syntactic complexity and memory-noun type. This, however, could be a result of the differences in the procedures used in our vs. Gordon et al.'s experiment. Specifically, in Gordon et al.'s study the comprehension questions were asked immediately after the sentences were read, whereas in our design, the reading of the sentences was followed by the memory recall task, which in turn was followed by the comprehension questions. The longer time lapse between the reading of the sentences and the comprehension questions in our design is likely to be responsible for overall lower comprehension accuracies (compared to Gordon et al.'s results), which could have potentially obscured some effects. Furthermore, by asking the participants to recall the memory items before answering the comprehension questions, we could have reduced the possible differential effect of memory-noun type on comprehension accuracies in subject- vs. object-extracted relative clause conditions, because after having attempted to select the memory-nouns from the set of available memory traces in memory, subjects might have better (although still quite poor, overall) access to the source-memory information—i.e., whether the noun came from the memory-list or from the sentence—about the sentence-nouns across the conditions.

Reading times

Because participants had to perform a memory task and to answer two comprehension questions for each sentence, the odds of getting all three correct were not very high overall (55%). As a result, we analyzed all trials, regardless of how the memory task was performed and how the comprehension questions were answered. The data patterns were very similar in analyses of smaller amounts of data, in which we analyzed (1) trials in which one or both of the comprehension questions were answered correctly, or (2) trials in which the memory task was performed perfectly. To adjust for differences in region lengths as well as overall differences in participants' reading rates, a regression equation predicting reading times from region length was derived for each participant, using all filler and target items (Ferreira & Clifton, 1986; Trueswell, Tanenhaus, & Garnsey, 1994, for discussion). For each region, the reading time predicted by the participant's regression equation was subtracted from the actual measured reading time to obtain a residual reading time. The statistical analyses of the raw reading time data produced the same numerical patterns (see Appendices C and D for tables of the raw and residual reading times). Reading time data points that were less than 100 ms in the raw data (indicating erroneous key presses) or more than three standard deviations away from the mean residual RT for a position within a condition were excluded from the analysis, affecting 2.2% of the data.

We computed a three-factor ANOVA crossing syntactic complexity (subject-extracted relative clause, object-extracted relative clause), memory load (one memory-noun, three memory-nouns) and memory-noun type (match, non-match) on the critical region (Region 2) consisting of the relative clause (“who consulted the cardiologist”/“who the cardiologist consulted”). The results are presented in Table 3. Importantly, the ANOVA revealed a three-way interaction among the three factors (marginal in the items analysis), such that syntactic complexity and the number of memory-nouns interacted in the conditions where the memory-nouns and the sentence-nouns were similar, but not in the conditions where they were dissimilar. This interaction is consistent with the idea that the amount of interference is a function of not only the representational characteristics of memory items but also the number of memory items.

In addition to the three-way interaction, we observed the following effects: (1) a main effect of syntactic complexity, (2) a main effect of memory-noun type, (3) an interaction between syntactic complexity and memory-noun type, and (4) a marginal interaction between memory-noun type and memory load. For comparisons between means of conditions, we report 95% confidence intervals (CIs) based on the mean squared errors of the relevant effects from the participant analyses (Masson &

Table 3
Analysis of variance results for reading times at the critical region

Source of variance	By participants			By items		minF'	
	df	$F1$ value	MSe	*df*	$F2$ value	*df*	minF'
2 × 2 × 2 ANOVA							
Synt complexity	1,43	48.47*	292,390	1,31	88.98*	1,73	31.3*
Memory load	1,43	1.11	486,815	1,31	2.40	1,71	<1
Memory-noun type	1,43	4.03	231,736	1,31	5.69*	1,74	2.35
Synt × load	1,43	<1	178,848	1,31	<1	1,63	<1
Synt × noun type	1,43	5.10*	183,822	1,31	6.37*	1,74	2.83
Load × noun type	1,43	8.07*	109,319	1,31	2.93	1,53	2.14
Synt × load × noun type	1,43	4.80*	123,237	1,31	3.19	1,65	1.91
2 × 2 ANOVA easy-load conditions							
Synt complexity	1,43	51.43*	151,563	1,31	92.05*	1,73	32.9*
Memory-noun type	1,43	<1	136,577	1,31	<1	1,31	<1
Synt × noun type	1,43	<1	777,01	1,31	<1	1,42	<1
2 × 2 ANOVA hard-load conditions							
Synt complexity	1,43	20.06*	319,676	1,31	35.90*	1,73	12.8*
Memory-noun type	1,43	8.88*	204,478	1,31	5.64*	1,64	3.44
Synt × noun type	1,43	6.58*	229,359	1,31	6.79*	1,72	3.34

Note. Significant effects are marked by asterisk.

Loftus, 2003). Syntactically more complex object-extracted relative clause conditions were read slower (278.2 ms) than subject-extracted relative clause conditions (−123.1 ms; 95% CI = 164.4 ms). The matching conditions (where the memory-nouns were occupations) were read slower (129.04 ms) than the non-matching conditions (26.06 ms; 95% CI = 146.4 ms). The effect of match affected only the more complex object-extracted relative clauses, but not the less complex subject-extracted relative clauses. Finally, the effect of match was only present in the hard-load conditions, but not in the easy-load conditions, although this interaction was marginal in the items analysis. Given that the three-way interaction revealed differences in the patterns of results between the easy-load and the hard-load conditions, we present the analyses for the easy-load and hard-load conditions separately. This will also allow us to compare our hard-load conditions, which are most similar to Gordon et al.'s (2002) design, to the four conditions of Gordon et al.'s experiment more easily.

We will now present the analyses for the easy-load conditions. Fig. 1 presents the mean residual RTs per region across the four easy-load conditions. A 2 × 2 ANOVA crossing syntactic complexity (subject-extracted relative clause, object-extracted relative clause), and memory-noun type (match, non-match) in the critical region revealed a main effect of syntactic complexity, such that the object-extracted relative clause conditions were read significantly slower (248.9 ms) than the subject-extracted relative clause conditions (−172.05 ms; 95% CI = 118.4 ms), but no other effects. The results are presented in Table 3. In the other three regions (Region 1, Region 3, and Region 4), there were no reliable effects, with the exception of an unpredicted interaction in Region 3, such that in the non-match conditions, the difference between the subject- and object-extracted relative clauses was larger than in the match conditions ($F1(1,43) = 4.07$; MSe = 424,429; $p < .05$; $F2(1,31) = 5.78$; MSe = 324,219; $p < .05$; min$F'(1,74) = 2.38$; $p = .13$). There is no reason to expect an interaction of this sort in this region, because there are no differences in the linguistic materials among the four conditions. Furthermore, the effect is not a spill-over effect from the previous region, because there is no trend for an interaction of this sort in Region 2. This effect is therefore likely to be spurious.

We will now present the analyses for the hard-load conditions. Fig. 2 presents the mean residual RTs per region across the four hard-load conditions. A 2 × 2 ANOVA crossing syntactic complexity (subject-extracted relative clause, object-extracted relative clause), and memory-noun type (match, non-match) in the critical region revealed two significant main effects and a significant interaction. First, the object-extracted relative clause conditions were read significantly slower (307.6 ms) than the subject-extracted relative clause conditions (−74.2 ms; 95% CI = 171.9 ms). Second, the match conditions were read significantly slower (218.2 ms) than the non-match conditions (15.2 ms; 95% CI = 137.5 ms). Finally, there was a significant interaction, such that the effect of match was only present in the object-extracted relative clause conditions and not in the subject-extracted relative clause conditions. The results are presented in Table 3. In the other three regions (Region 1, Region 3, and Region 4), the only reliable effect was that of syntactic complexity in Region

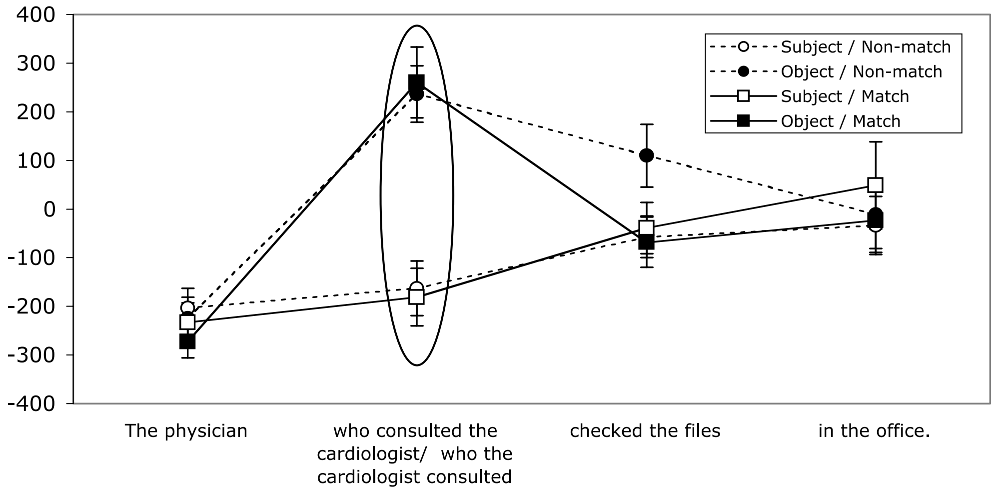

Fig. 1. Reading times per region in the four easy-load (one memory-noun) conditions. Error bars indicate standard errors. The critical region is circled.

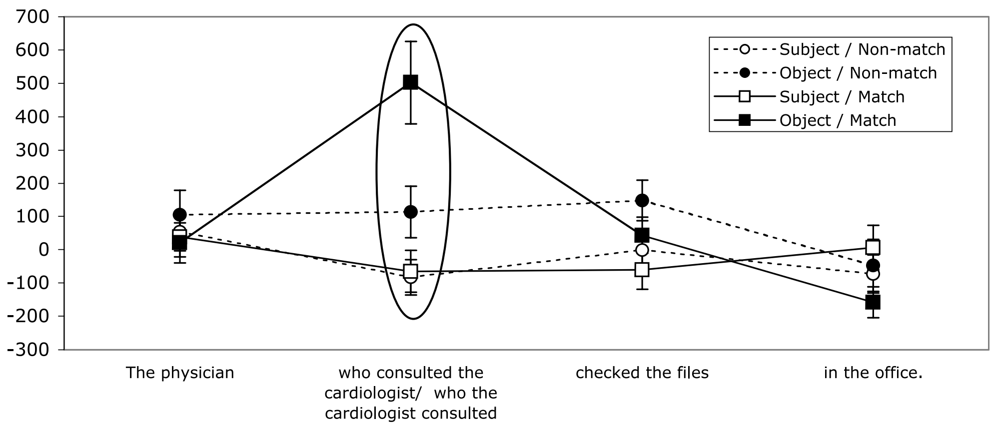

Fig. 2. Reading times per region in the four hard-load (three memory-nouns) conditions. Error bars indicate standard errors. The critical region is circled.

3, such that object-extracted relative clause conditions were read slower (95.5 ms) than subject-extracted relative clause conditions (−30.8 ms; 95% CI = 113.2 ms) ($F1(1,43) = 5.07$; MSe = 702,610; $p < .05$; $F2(1,31) = 6.56$; MSe = 462,948; $p < .02$; $\min F'(1,74) = 2.85$; $p = .095$). This effect could be a possible spillover effect from Region 2.

General discussion

The most interesting result presented here is an interaction between syntactic complexity and the memory-noun/sentence-noun similarity during the critical region of the linguistic materials in the hard-load (three memory-nouns) conditions: people processed object-extracted relative clauses more slowly when they had to maintain a set of nouns that were similar to the nouns in the sentence than when they had to maintain a set of nouns that were dissimilar from the nouns in the sentence; in contrast, for the less complex subject-extracted relative clauses, there was no reading time difference between the similar and dissimilar memory load conditions. In the easy-load (one memory-noun) conditions, no interaction between syntactic complexity and memory-noun/sentence-noun similarity was observed.

These results provide evidence against the hypothesis whereby there is a pool of domain-specific verbal working memory resources for sentence comprehension, contra Caplan and Waters (1999). Specifically, the results reported here extend the off-line results reported by Gordon et al. (2002) who—using a similar dual-task paradigm—provided evidence of an interaction between syntactic complexity and memory-noun/

sentence-noun similarity in response-accuracies to questions about the content of the sentences. Although there was also a suggestion of an on-line reading time interaction in Gordon et al.'s experiment, this effect did not reach significance. The effect of memory-noun type in subject-extracted conditions was 12.7 ms per word (50.8 ms over the four-word relative clause region), and the effect of memory-noun type in object-extracted conditions was 40.1 ms (160.4 ms over the four-word relative clause region). (Thanks to Bill Levine for providing these means). In the hard-load conditions of our experiment, the effect of memory-noun type in subject-extracted conditions over the four-word relative clause region is 12 ms in raw RTs (18 ms in residual RTs), and the effect of memory-noun type in object-extracted conditions is 324 ms in raw RTs (388 ms in residual RTs). This difference in effect sizes of memory-noun type between Gordon et al.'s and our experiment is plausibly responsible for the interaction observed in the current experiment, and the lack of such an interaction in Gordon et al.'s study. The current results therefore demonstrate that Gordon et al.'s results extend to on-line processing.

Furthermore, the interaction among the three factors provides evidence in support of the hypothesis whereby the amount of interference is a function of both the representational characteristics of memory items, and the number of memory items. Specifically, syntactic complexity interacted with memory load (manipulated in terms of the number of items) only in the context of similar elements. In light of these results, it is possible to explain the lack of interactions between syntactic complexity and number of memory items in the previous dual-task experiments, which used digit-span as a secondary verbal working memory task (e.g., Waters et al., 1987, Waters, Caplan, & Rochon, 1995). Because digits are qualitatively very different from the nouns in the sentence materials (similar to the non-match conditions in our experiment), there is no reason to expect that the memory load should interact with syntactic complexity.

To account for the observed on-line interaction between syntactic complexity and the complexity of a non-linguistic verbally-mediated task, we would like to elaborate Gordon et al.'s proposal. One possible way to spell out the interaction between syntactic complexity and the memory-noun/sentence-noun similarity is in terms of local vs. non-local integrations, as suggested by Gordon et al. (p. 426). Gibson (1998, 2000) and Grodner and Gibson (2005) provide a framework to formalize this idea. The integration of a new word can either be local—to the preceding word in the sentence—or non-local, requiring a retrieval of a word further back in the input stream. In the case of a local integration, the incoming word can be immediately connected to the preceding word, because there are no intervening potential attachment sites. Thus, in subject-extracted relative clauses—at the point of integrating the embedded verb with the embedded subject—no search for an attachment site is needed, and the existence of a set of stored memory items will probably not cause any interference in forming the connection. However, when the integration is non-local, the target attachment site needs to be retrieved from memory, and there may be interference from the intervening potential attachment sites (e.g., McElree, Foraker, & Dyer, 2003). Specifically, in object-extracted relative clauses, an embedded subject intervenes between the subject of the main clause and the embedded verb, and therefore may cause interference. Furthermore, when there is a list of stored items in working memory, the similarity of these items to the target attachment site (in this case, the subject of the main clause, which is also the object of the embedded verb) may further increase the difficulty of the integration. Thus, there should be some difficulty with both non-local integrations, but more difficulty in the condition where the memory items are similar to the item that needs to be retrieved from memory. The interaction between syntactic complexity and similarity between memory-nouns and sentence-nouns can therefore be accounted for in terms of integration difficulty: (a) there is little difficulty for *local* integrations, with either similar or dissimilar memory items; and (b) for *non-local* integrations, the condition with similar memory items is much more difficult to process than the condition with dissimilar memory items.

Although we have discussed the observed effects in terms of difficulty at the retrieval stage of the memory process, in theory it is possible that the effects occur during the encoding and/or during the maintenance stage of the memory process. For example, it is possible that the encoding process becomes increasingly more costly with an increase in the number of shared features between the to-be-remembered items. Specifically, when the sentence-nouns are being encoded into memory, the difficulty of this process may be a function of the similarity of the nouns that have been recently encoded and are currently stored in the memory store. Similarly, it is possible that the difficulty of the storage process increases with an increase in the overlap in representational characteristics of the to-be-remembered items.

Research on short-term memory provides some suggestive evidence in favor of the hypothesis whereby proactive interference effects take place at the retrieval stage (Gardiner, Craik, & Birtwistle, 1972; as cited in Crowder, 1976). Gardiner et al. (1972; as cited in Crowder, 1976) conducted an experiment in order to identify the locus of proactive interference effects among the three different stages of the memory process, and provided evidence for placing these effects at the stage of retrieval. Specifically, Gardiner et al. conducted an experiment where subjects were presented with four lists, and tested for their memory of the items in the fourth list. The items in the first three lists all came from the same semantic category (e.g., garden flowers). The items in the fourth list came from a slightly different semantic category (e.g., wild flowers). Any version of a proactive interference hypothesis predicts interference effects from the first three lists on the memory for the items from the fourth list. In the attempt to identify the locus of the proactive interference effects, Gardiner et al. divided the subjects into three groups. The control group was not informed with regard to the difference in categories between the first three lists and the fourth list; the two experimental groups were informed about the difference in categories either before, or after the presentation of the fourth list. If the proactive interference effects took place at the stage of encoding or storage, then the group that was informed about the difference in categories between the first three lists and the fourth list *after* the presentation of the fourth list should behave similarly to the control group, as they would not be able to take advantage of the distinctiveness of the items in the fourth list until the retrieval stage. In contrast, if the proactive interference effects took place at the stage of retrieval, then this experimental group should behave similarly to the group that was informed about the difference in categories between the first three lists and the fourth list *before* the presentation. The results were consistent with the hypothesis whereby the interference effects take place at the retrieval stage. Thus, based on evidence from the short-term memory research, and given the fact that it is easiest to conceptualize the interference effects we observed in terms of relative ease/difficulty of retrieval in local/non-local integrations, we tentatively conclude that the locus of proactive interference effects in sentence processing is at the retrieval stage of the memory process.

Acknowledgments

We thank David Badre, Mara Breen, David Caplan, Neal Pearlmutter, Adam Szczegielniak, Gloria Waters, three anonymous reviewers, and the audience at CUNY 2005. We are especially grateful to Timmy Desmet for his patience in reading multiple drafts of the paper.

Appendix A. Linguistic materials

One of the four subject-/object-extracted relative clause versions is shown below for each of the 32 items. The other three versions can be generated as exemplified in (1) below.

(1) a. Subject-extracted, version 1:
The physician who consulted the cardiologist checked the files in his office.
b. Subject-extracted, version 2:
The cardiologist who consulted the physician checked the files in his office.
c. Object-extracted, version 1:
The physician who the cardiologist consulted checked the files in his office.
d. Object-extracted, version 2:
The cardiologist who the physician consulted checked the files in his office.

(2) The babysitter who liked the parents planned a trip to Puerto Rico.
(3) The banker who informed the chairman invested a million in a start-up.
(4) The violinist who flattered the cellist played a piece from the symphony.
(5) The burglar who wounded the policeman reloaded the revolver in a hurry.
(6) The carpenter who punched the electrician quit the job a week later.
(7) The accountant who advised the statistician calculated the costs of the project.
(8) The model who approached the artist signed the contract for a year.
(9) The student who trusted the professor answered the question about the experiment.
(10) The mobster who attacked the dealer organized some crimes in New York.
(11) The investigator who overheard the cop closed the case without an arrest.
(12) The actor who respected the starlet forgot the lines during the scene.
(13) The defendant who misled the lawyer blamed the system for the conviction.
(14) The count who adored the princess brought a gift to the reception.
(15) The bachelor who pursued the socialite owned a company in the area.
(16) The councilman who kissed the secretary covered the expenses for the party.
(17) The contestant who offended the host ruined the show for the audience.
(18) The mathematician who addressed the physicist offered the proof at the conference.
(19) The diplomat who insulted the congressman ended the negotiations on the spot.
(20) The priest who thanked the nun founded the shelter near the church.
(21) The analyst who queried the governor proposed some changes to the plan.
(22) The farmer who questioned the expert promoted the product at the fair.
(23) The official who harassed the manager questioned the policy of lowering wages.
(24) The clerk who disliked the director typed the letter to the administration.
(25) The guitarist who recommended the band recorded the song for the album.
(26) The salesman who resented the cashier mislabeled the products in the brochure.
(27) The waiter who invited the cook tasted the sauce for the meat.
(28) The medic who assisted the doctor borrowed the instrument for the surgery.

(29) The passenger who befriended the stewardess remembered the flight across the Atlantic.
(30) The cheerleader who bothered the quarterback attended the game at the college.
(31) The animator who criticized the producer offered a solution to the problem.
(32) The dictator who despised the dissident gave a speech about the protests.

Appendix B. Memory-nouns statistics

	Lexical frequency[a]	Length (in syllables)
Occupation memory-nouns	10,717	2.18
Names memory-nouns	10,715	2.24
T test	.98	.16

[a] Lexical frequencies were matched using a Usenet corpus of 1.2 billion words.

Appendix C. Residual reading times

Residual reading times in milliseconds, as a function of syntactic complexity, memory load, and memory-noun type (standard errors in parentheses) for Region 1

Memory load	Syntactic complexity			
	Subject-extraction (easy)		Object-extraction (hard)	
Memory-noun	Match	Non-match	Match	Non-match
One noun (easy)	−233 (40)	−204 (40)	−273 (33)	−225 (44)
Three nouns (hard)	38.6 (61)	54.03 (57)	20.7 (60)	106 (73)

Residual reading times in milliseconds, as a function of syntactic complexity, memory load, and memory-noun type (standard errors in parentheses) for Region 2

Memory load	Syntactic complexity			
	Subject-extraction (easy)		Object-extraction (hard)	
Memory-noun	Match	Non-match	Match	Non-match
One noun (easy)	−181 (59)	−163 (56)	261 (73)	237 (58)
Three nouns (hard)	−65.2 (63)	−83.1 (53)	502 (123)	113 (78)

Residual reading times in milliseconds, as a function of syntactic complexity, memory load, and memory-noun type (standard errors in parentheses) for Region 3

Memory load	Syntactic complexity			
	Subject-extraction (easy)		Object-extraction (hard)	
Memory-noun	Match	Non-match	Match	Non-match
One noun (easy)	−38.8 (53)	−56.9 (43)	−68.02 (52)	110 (65)
Three nouns (hard)	−61.2 (58)	−0.47 (57)	43.4 (54)	148 (61)

Residual reading times in milliseconds, as a function of syntactic complexity, memory load, and memory-noun type (standard errors in parentheses) for Region 4

Memory load	Syntactic complexity			
	Subject-extraction (easy)		Object-extraction (hard)	
Memory-noun	Match	Non-match	Match	Non-match
One noun (easy)	49.1 (89)	−33.7 (59)	−23.1 (67)	−10.9 (71)
Three nouns (hard)	6.48 (68)	−72.6 (57)	−159 (46)	−47.1 (78)

Appendix D. Raw reading times

Raw reading times in milliseconds, as a function of syntactic complexity, memory load, and memory-noun type (standard errors in parentheses) for Region 1

Memory load	Syntactic complexity			
	Subject-extraction (easy)		Object-extraction (hard)	
Memory-noun	Match	Non-match	Match	Non-match
One noun (easy)	771 (51)	783 (49)	734 (46)	769 (41)
Three nouns (hard)	996 (81)	1024 (99)	986 (94)	1085 (99)

Raw reading times in milliseconds, as a function of syntactic complexity, memory load, and memory-noun type (standard errors in parentheses) for Region 2

Memory load	Syntactic complexity			
	Subject-extraction (easy)		Object-extraction (hard)	
Memory-noun	Match	Non-match	Match	Non-match
One noun (easy)	1297 (77)	1304 (75)	1720 (118)	1729 (103)
Three nouns (hard)	1445 (104)	1457 (108)	1985 (180)	1661 (128)

Raw reading times in milliseconds, as a function of syntactic complexity, memory load, and memory-noun type (standard errors in parentheses) for Region 3

Memory load	Syntactic complexity			
	Subject-extraction (easy)		Object-extraction (hard)	
Memory-noun	Match	Non-match	Match	Non-match
One noun (easy)	1172 (52)	1171 (59)	1147 (73)	1326 (102)
Three nouns (hard)	1116 (63)	1203 (82)	1269 (84)	1379 (88)

Raw reading times in milliseconds, as a function of syntactic complexity, memory load, and memory-noun type (standard errors in parentheses) for Region 4

Memory load	Syntactic complexity			
	Subject-extraction (easy)		Object-extraction (hard)	
Memory-noun	Match	Non-match	Match	Non-match
One noun (easy)	1142 (92)	1062 (70)	1089 (86)	1113 (75)
Three nouns (hard)	1137 (93)	1053 (87)	974 (71)	1072 (101)

References

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.). *The psychology of learning and motivation* (Vol. 8, pp. 47–89). New York: Academic Press.

Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.

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

Chomsky, N. (1986). *Knowledge of language: its nature, origin, and use*. New York: Praeger Publishers.

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

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

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

Fodor, J. (1983). *The modularity of mind: an essay on faculty psychology*. Cambridge, MA: MIT Press.

Gardiner, J. M., Craik, F. I. M., & Birtwistle, 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 Y. Miyashita, A. Marantz, & W. O'Neil (Eds.), *Image, language, brain*. Cambridge, MA: MIT Press.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*, 1–67.

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

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input. *Cognitive Science, Vol. 29*(2), 261–290.

Hanley, J. R., Young, A., & Pearson, N. A. (1991). Impairment of the visuo-spatial scratchpad. *Quarterly Journal of Experimental Psychology, 43A*, 101–125.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93*, 411–428.

Jonides, J., Smith, E. E., Koeppe, R. A., Awh, E., Minoshima, S., & Mintun, M. A. (1993). Spatial working memory in humans as revealed by PET. *Nature, 363*, 623–625.

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

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processing in reading comprehension. *Journal of Experimental Psychology: General, 111*, 228–238.

Kanwisher, N., McDermott, J., & Chun, M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for the perception of faces. *Journal of Neuroscience, 17*, 4302–4311.

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

Lewis, R. (1996). A theory of grammatical but unacceptable embeddings. *Journal of Psycholinguistic Research, 25*, 93–116.

Masson, M. E., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology, 57*, 202–220.

McDermott, J., & Hauser, M. (in press). The origins of music: Innateness, development, and evolution. *Music Perception*.

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

Peretz, I., & Hyde, K. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Science, 8*(7), 362–367.

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

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125*(1), 4–27.

Vallar, G., & Shallice, T. (Eds.). (1990). *Neuropsychological impairments of short-term memory*. New York: Cambridge University Press.

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

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

Waters, G., Caplan, D., & Hildebrandt, N. (1987). Working memory and written sentence comprehension. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading*. Erlbaum.

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