Journal of Psycholinguistic Research, Vol. 31, No. 6, November 2002 (© 2002)

Individual Differences in Sentence Memory

Rose Roberts[1] and Edward Gibson[1,2]

Results from an experiment with two parts are presented in this paper. In part one, participants listened to sentences containing two, three, four, or five clauses, and were asked questions about the content of the sentences. The results of part one demonstrate that an important unit of representation in sentence memory is the clause, and not some other component of discourse structure. In part two, the same group of participants performed eight different short-term storage/working memory tasks. A composite complex span score was computed for each participant based on three working memory tasks closely based on Daneman & Carpenter's (1980) reading span task. This working memory measure was significantly correlated with the participants' performance on the sentence memory task in part one. A second working memory measure—N-back—was also significantly correlated with the participants' performance on the sentence memory task, and there was no correlation between their performance on the complex span task and the N-back task. It is therefore concluded that (i) working memory consists of a number of dissociable components; and (ii) memory for sentences taps into more than one of these working memory components. Furthermore, the high correlations of sentence memory with the complex span and the N-back tasks (neither of which are language processing tasks) suggests that memory for sentences is not simply a result of linguistic experience; rather, it is likely that an independent working memory component contributes to participants' performance on the sentence memory task.

KEY WORDS: Working memory; sentence comprehension; sentence memory; reading span.

INTRODUCTION

The experiment presented here addresses the question of the relevant memory unit in understanding and remembering sentences, and examines the relation between working memory and memory for sentences. Caplan and Waters (1999) distinguish two types of linguistic processing: interpretive processing, the assignment of syntactic and semantic structure to sentences;

[1] Massachusetts Institute of Technology, Cambridge, Massachusetts.
[2] To whom all correspondence should be addressed. Department of Brain and Cognitive Sciences, NE20-459, MIT, Cambridge MA 02139. Phone: (617) 253-8609; Fax: (617) 258-8654; email: gibson@psyche.mit.edu

0090-6905/02/1100-0573/0

and post-interpretive processing, the use of interpreted sentences for other verbal tasks, such as answering questions, reasoning, and planning actions. Historically, the clause has been hypothesized to be an important storage unit in on-line sentence comprehension (interpretive processing) (Kimball, 1973) and in off-line sentence memory (post-interpretive processing) (Blauberg & Braine, 1974; Jarvella, 1971; see Tanenhaus & Trueswell, 1995, for an overview). However, recent evidence examining interpretive processing of sentences has challenged this idea, and has suggested that incomplete dependencies and discourse structures (e.g., discourse referents such as objects and events), rather than clauses, are the relevant units for interpretive processing (Gibson, 1998, 2000). It would therefore be appropriate to reanalyze the data underlying the hypothesis that the clause is the unit of post-interpretive processing. A review of the literature shows that the data used to support this hypothesis are open to alternative interpretations: Specifically, there are a number of confounds in previous studies, and consequently the data do not unambiguously support the claim that the clause is the unit of sentence memory. The experiment reported here was designed to determine whether the clause is the unit of sentence memory, or whether, as in interpretive processing, sentence memory might be a function of the number of new discourse referents or content words in the sentence.

In measures of on-line and off-line sentence processing, researchers have observed large individual differences in performance (speed and/or accuracy of responses). The most common explanation for this variance is in terms of individual differences in working memory, the ability to store and manipulate information used in complex cognitive tasks (Baddeley, 1983; Daneman & Carpenter, 1980; Just & Carpenter, 1992). An alternative view, recently proposed by MacDonald and Christiansen (2002), is that working memory used for language processing does not exist: Individual differences are due to differences in reading skill and experience with language. The second goal of this study was to determine whether individual differences in working memory, independent of reading skill, can explain variance in sentence memory.

STORAGE UNITS IN ON-LINE SENTENCE COMPREHENSION

An early proposal regarding the storage unit in on-line sentence comprehension was made by Kimball (1973). He proposed that sentence comprehension was clause-based, such that at most two partially processed clauses could be maintained in working memory at one time. This proposal accounted for the difficulty associated with processing *nested* structures such as (1):

(1) [s The student [who [s the professor [who [s the scientist collaborated with]] had advised]] copied the article].

A syntactic category A is said to be nested (or *center embedded*) within another category B if B contains A, a constituent to the left of A, and a constituent to the right of A. In (1), the relative clause (RC) "who the professor . . . had advised" is nested within the sentence "the student . . . copied the article." Furthermore, a second RC "who the scientist collaborated with" is nested within the first embedded sentence "the professor . . . had advised." By contrast, left- or right-branching structures are much easier to understand than nested structures. For example, the right-branching structure in (2) has the same meaning as its nested counterpart in (1) at the level of thematic structure, but it is much easier to understand:

(2) The scientist collaborated with the professor who had advised the student who copied the article.

Kimball's clause-based proposals accounts for the contrast between nested and right-branching RCs as follows. Processing the first subject "the student" in (1) causes the initiation of a new clause that will be completed when the verb "copied" and its immediate dependents are located in the input string. Processing the following two subjects of the embedded RCs ("the professor" and "the scientist") causes the initiation of two further clauses, resulting in a total of three partially processed clauses, which is more than the proposed resource capacity. By contrast, there is never more than one incomplete clause while processing the right-branching structure in (2), so this sentence is processed without difficulty.

Although the clause-based proposal accounted for some nesting complexity effects, recent results suggest that the difficulty that people have in processing nested structures is not because of a clause-based processing mechanism. Rather, the difficulty that people have with processing nested structures seems to depend on two factors: (i) the distance between the syntactic dependents in a sentence, and (ii) the number of syntactic heads that are required to form a grammatical sentence from a partially processed input string (Gibson, 1998, 2000; for an alternative, see Lewis, 1996). According to the theory proposed by Gibson and colleagues, the point of maximal complexity in processing the sentence in (1) occurs at the point of processing the second verbal complex "had advised". There are two complex integrations to be performed at this point. First, the NP "the professor" must be integrated as the subject of the verb "advised," assigning the agent thematic role in the process. Second, the pronoun "who," which is coindexed with the sentence-initial NP "the student" must be integrated as object of the verb "advised" by coindexation with an empty category in object position of this verb. Neither of these integrations is local. The subject-verb integration crosses the intervening relative clause "who the scientist collaborated with." The pronoun-object-gap integration is even longer distance. In addition, at the point when these integrations are being made, three syntactic heads must be stored to ensure the formation of a grammatical sentence: (i) the top-level verb, (ii) a verb to head the first RC, and (iii) an NP empty-category to be coindexed with the first RC pronoun "who." As a result, the two integrations are very difficult for people to make.

Empirical evidence that the distance metric is not clause-based is provided by Gibson (1998). Evidence that the distance metric is based on the discourse accessibility of the material in the interim is provided by Warren and Gibson (2002), who provide results from both an off-line task and an on-line task. For example, Warren and Gibson (2002) present questionnaire results that demonstrate that doubly nested RC structures are easier to process when a first- or second-person pronoun is in the subject position of the most embedded RC, as in (3), as compared with a similar structure in which the NP in the most embedded position (e.g., "the scientist" in [1]) introduces a new object into the discourse (Bever, 1970; Kac, 1981):

(3) The student who the professor who I collaborated with had advised copied the article.

The integrations at the predicates "collaborated with," "had advised" and "copied" all cross the most embedded subject in (3) and (1). If integration difficulty depends on the complexity of constructing or accessing the discourse structures for the words in the interim, these integrations are all easier in (3) than in (1). In particular, more effort is needed to construct a new discourse object for the NP "the scientist" in (1) than to access the highly accessible referent for "I" in (3). This difference in discourse complexity results in easier integration at each of the three predicates in (3) than in (1). As a result, (3) is easier to process than (1). Note that a clause-based proposal does not predict this complexity contrast, because there are three partially processed sentences at the most embedded subject of (3), just as in (1).

Given the shift in recent theory from an interpretive processing mechanism based on clause units to one based on linear distance in terms of discourse structure, it is worthwhile to investigate whether the evidence for clause-based post-interpretive processing might be reinterpreted as well.

UNITS OF SENTENCE MEMORY

Clauses, and not within-clause phrase boundaries, appear to be the units of segmentation during speech perception. Clicks occurring during the auditory presentation of sentences are misheard to occur between clause boundaries but not between phrase boundaries or between individual words (Bever, Lackner, & Kirk, 1969). Controlling for lexical items and serial position, words occurring in recent clauses are recognized more quickly than words occurring in early clauses (Caplan, 1972; Chang, 1980). Early research examining the units and capacity of sentence memory appears to show that the clause is also a unit of sentence memory. Two pieces of evidence support this assumption: First, memory declines as a function of the number of clauses in the sentences (Blauberg & Braine, 1974). Second, verbatim recall is highly accurate for the most recent clause heard, but not for earlier clauses (Jarvella, 1971).

However, there are a number of other candidate units of sentence memory: Memory could be a function of the number of words in the sentence, the number of content words (nouns, verbs, adjectives, etc.), or the number of new discourse objects introduced into the mental model. Glanzer and Razel (1974) showed that short, familiar sentences (proverbs) could be held as units in short-term memory, but that unfamiliar sentences were recalled less accurately, suggesting that not all clauses can be held as chunks in short-term storage. Given the ample experimental evidence that meaningful word strings are recalled more accurately than unrelated lists (Gershberg & Shimamura, 1995; Larkin & Burns, 1977; Miller, 1956; Miller & Isard, 1964), it is unlikely that memory is simply a function of the number of words in the sentence. However, the studies reviewed above do not rule out the other alternatives. Because of confounds in these studies, the results do not unambiguously support the hypothesis that clauses are the units of sentence memory.

In Blauberg and Braine's (1974) study, participants were presented with 30 sentences, 6 at each length from 2 to 6 clauses. They were then presented with a probe, one of the clauses from the sentence with the subject or object noun missing, and asked to produce the appropriate noun. Recall accuracy declined as a function of the number of clauses in the sentence, with an average of 6/6 probe questions answered correctly for 2-clause sentences, and 3/6 probes answered correctly for 6-clause sentences. Although these results show that memory declines as a function of the number of clauses in the sentence, they do not show that the clause is the unit of sentence memory. Sentences with more clauses were longer, and thus it is impossible to determine whether the difficulty with long sentences was due to their greater length or their greater number of clauses. Determining whether clauses are the units of sentence memory requires a comparison of sentences in which the effects of the number of words and the number of clauses can be separated.

In Jarvella's (1971) study, experiment participants were played a connected discourse that was periodically interrupted, with participants instructed to recall as much of the previous material as possible. He examined recall accuracy for the most recently heard two (Experiment 2) or three (Experiment 1) clauses. Overall, he found that the most recent clause was recalled highly accurately, with earlier clauses recalled less accurately, results that appeared to support the idea of clause-based processing. However, there are two problems with the experimental materials and methods that preclude one from interpreting these results as support for clause-based processing. First, the results of Experiment 2 showed that although there was an overall effect of clause position (early versus recent), when the two clauses were within the same sentence, recall was still accurate for the earlier clause (early clause = .84, recent clause = .97), whereas when they were in different sentences, recall was more accurate in the most recent clause (.95) than in the earlier clause (.63). This result provides more support for the idea that sentences are the units of memory, rather than clauses. But problems with the materials render even this interpretation problematic: When the recent and early clauses were in different sentences, the sentence containing the early clause was more complex than when the recent and early clauses were in the same sentence. Sample sentences from Jarvella (1971) illustrate this problem:

(4) Early clause in same sentence:
He and the others were labeled as Communists. McDonald and his top advisors hoped this would keep Rarick off the ballot.

(5) Early clause in previous sentence:
That he could be intimidated was what McDonald and his top advisors hoped. This would keep Rarick off the ballot.

Whereas the words in the final two clauses ([MacDonald and his top advisors hoped] [this would keep Rarick off the ballot]) are identical in the two conditions, the sentence structure in (5) is more complex. The first sentence in (5) contains the cleft sentential subject "that he could be intimidated," which is harder to understand than the simpler subject-verb-object sequence in (4) (Frazier & Rayner, 1988; Gibson, 1998). This factor would render the sentences in (4) harder to recall verbatim, as was required in this study. Measuring verbatim recall is the second weakness of this experiment: Requiring participants to recall sentences verbatim focuses on the surface structure of sentences, rather than the conceptual or propositional content. Potter and colleagues have provided evidence that verbatim recall, relying on a briefly held phonological record, can be dissociated from memory for the conceptual content of sentences, memory that retains an abstract representation of propositional content without retaining the exact lexical or syntactic form of the sentence (Lombardi & Potter, 1992; Potter, 1993; Potter & Lombardi, 1990; Potter, Moryadas, Abrams, & Noel, 1993; Potter, Valian, & Faulconer, 1977). Even if the experiments in Jarvella (1971) had succeeded in showing that clauses are the units of verbatim recall, this finding could not be used as evidence that clauses are the units of sentence memory when sentence content is probed rather than surface form and lexical items.

Thus, although the evidence that the clause is the unit of segmentation during speech perception is convincing (Bever *et al.*, 1969; Caplan, 1972; Chang, 1980), neither Blauberg and Braine (1974) nor Jarvella (1971) provide conclusive evidence that the clause is the unit of sentence memory. Given the evidence that on-line sentence comprehension is not clause-based, it is therefore worthwhile to re-evaluate the hypothesis that sentence memory is clause-based while controlling for the confounds in earlier experiments observed here.

THE ROLE OF WORKING MEMORY IN MEMORY FOR SENTENCES

The most influential model of working memory is the multicomponent (MC) model proposed by Baddeley and colleagues (Baddeley, 1983, 1996; Baddeley & Hitch, 1994), which has been extended and modified by other researchers (Lehto, 1996; Martin & Romani, 1994; Smith & Jonides, 1997). In MC models, the verbal part of working memory consists of at least two components: The short-term store (STS), used for storing and rehearsing verbal information using a phonological code[3] (Awh *et al.*, 1996; Baddeley, 1996; Basso, Spinnler, Vallar, & Zanobio, 1982; Fiez *et al.*, 1996; Paulesu, Frith, & Frackowiak, 1993; Vallar, Betta, & Silveri, 1997); and the central executive (CE), used for allocating attention, planning, inhibiting nonrelevant responses, and coordinating resources demanded by concurrent tasks (Baddeley, 1996; D'Esposito *et al.*, 1995; Lehto, 1996). Remembering sentences and answering questions about them could depend primarily on the STS, the CE, or both. One goal of the current experiments is to determine whether the STS and the CE make independent contributions to sentence memory capacity.

Gathercole and Baddeley (1993) have hypothesized that the STS and the CE make dissociable contributions to language understanding. For sentence processing, the STS is hypothesized to maintain a phonological record of sentences just heard or read. Such a record would be particularly useful for sentences that are initially understood incorrectly and must be reanalyzed (so-called "garden-path" sentences such as "Tom told the children the story scared a riddle" [Frazier, 1978]), or that contain a long list of items

[3] In Baddeley's model, passive verbal storage includes two components: the phonological store, used for holding information, and the articulatory loop, used to rehearse information using a speech-based code (Baddeley, 1983; Baddeley, Vallar, & Wilson, 1987). Such a fine-grained analysis is not relevant to the hypotheses under investigation here, so for the sake of simplicity we are including both components in the STS.

to be remembered (e.g., “Please go to the store and buy bread, milk, eggs, cheese, oranges, and spinach.”). Evidence for the role of the STS in sentence processing includes the finding that participants are impaired at comprehending long, complex sentences when they have to concurrently articulate irrelevant words (Baddeley, Eldridge, & Lewis, 1981).

An early idea about the role of STS in language comprehension held that sentence comprehension requires an ordered, verbatim representation of the words just heard. However, this idea was contradicted by the discovery that patients with a severe deficit in the STS (digit or word spans of one to four items) are relatively unimpaired in understanding language (Baddeley *et al.*, 1987; Basso *et al.*, 1982; Belleville, Peretz, & Arguin, 1992; Martin, 1987, 1993; McCarthy & Warrington, 1987a, 1987b; Saffran & Marin, 1975; Vallar *et al.*, 1997). When tested in detail, such patients show impairment only in understanding very long or complex sentences (see Caplan & Waters, 1990, for a review). Other results that questioned the relevance of a verbatim representation for sentence comprehension come from Potter and colleagues, who found that people quickly form conceptual representations of sentence meaning while losing information about the exact words and syntactic structure (Lombardi & Potter, 1992; Potter, 1993; Potter & Lombardi, 1990; Potter *et al.*, 1993; Potter *et al.*, 1977). Such evidence indicates that the STS is not crucial for first-pass comprehension, although it may be useful for higher-level linguistic interpretations that lag behind on-line comprehension processes.

According to Gathercole and Baddeley, the CE is used for syntactic and semantic processing and for storing the intermediate and final products of such processing during performance of post-interpretive tasks such as answering questions, verifying the truth of statements, and reasoning from given propositions. The CE also has been claimed to be important for integrating new propositions with the representation of a text and maintaining the predicate-argument structure of propositions (Caplan & Waters, 1999; Gathercole & Baddeley, 1993; Kintsch & van Dijk, 1978). Evidence for the role of the CE in sentence memory includes the fact that tasks like Daneman & Carpenter’s (1980) Reading Span, which require the coordination of storage and processing (a CE function), correlate more highly with measures of reading comprehension (i.e., verbal SAT, answering factual questions about passage, or understanding pronoun referents) than do measures of STS alone (i.e., digit or word span). Another source of evidence about the role of the CE in sentence memory comes from examining the performance of patients with impaired CE function. Patients with Alzheimer’s disease (AD) have severe impairments in CE functions (Baddeley, Logie, Bressi, Sala, & Spinnler, 1986; Baddeley, Bressi, Sala, Logie, & Spinnler, 1991). Waters and colleagues (Waters & Caplan, 1997; Waters, Caplan, &

Rochon, 1995) have presented evidence that patients with AD are impaired in post-interpretive processing of sentences with more than one proposition, and are more disrupted than control participants under dual-task conditions, which require the coordinating function of the CE. These results that the CE may be crucial for normal post-interpretive sentence processing, which involves sentence memory.

MEASURING WORKING MEMORY

Any researcher attempting to measure working memory capacity faces a serious challenge in deciding what test to use. A review of the working memory literature shows that several very different tasks are commonly used to assess working memory (Baddeley, Logie, Nimmo-Smith, & Brereton, 1985; Braver *et al.,* 1997; Case, Kurland, & Goldberg, 1982; Daneman & Carpenter, 1980; Daneman & Merikle, 1996; Dobbs & Rule, 1989; Engle, Cantor & Carullo, 1992; Klapp, Marshburn, & Lester, 1983; LaPointe & Engle, 1990; Petrides, Alivisatos, Meyer, & Evans, 1993; Salthouse, 1990; Turner & Engle, 1989, Waters & Caplan, 1996). Several recent studies, however, have suggested that scores on commonly used working memory tests may not in fact be highly correlated, and that the different tests may be differentially sensitive to processes such as storage, response time, rapid stimulus manipulation, and other CE functions (Dobbs & Rule, 1989; Lehto, 1996; Waters & Caplan, 1996). Roberts and Corkin (1997) compared performance on several different tests of working memory, and finding that they were not highly intercorrelated and that scores on different tests were predicted by measures of different component variables (short-term storage and processing speed). These results suggest that different working memory tests are sensitive to different components of working memory.

The relation between working memory and language comprehension or other cognitive functions has been assessed using two different analytical techniques: correlational studies and the individual differences approach. In correlational studies (Babcock & Salthouse, 1990; Case *et al.,* 1982; Daneman & Carpenter, 1980; Daneman & Carpenter, 1983; Daneman & Merikle, 1996; Engle *et al.,* 1992; LaPointe & Engle, 1990; Turner & Engle, 1989; Waters & Caplan, 1996), participants are given a variety of tests of working memory and other cognitive functions, and correlations, multiple regressions, and factor analyses are used to determine what variables are related to the cognitive function of interest. These practices allow researchers to determine relations among large numbers of variables, and the extent to which different tests contribute common and unique variance to the measure of interest (Engle *et al.,* 1992). One limitation of correlational approaches is that multiple comparisons require large numbers of participants in order to be reliable. Another problem in the literature (although not inherent in the approach) is that with large numbers of participants, statistically significant correlations may account for only a small amount of the variance on a given test. For example, with a large number of participants, an r of .25 may be statistically significant at the $p < .05$ level, but would explain less than 10% of the variance. Such a small correlation would not constitute a sufficient explanation of the relation between the correlated variables.

The second approach to studying the relation between working memory and language is the individual differences approach (Just & Carpenter, 1992; Just, Carpenter, & Keller, 1996; King & Just, 1991; King & Kutas, 1995; MacDonald, Just, & Carpenter, 1992; Miyake, Carpenter, & Just, 1994; Miyake, Carpenter, & Just, 1995; Miyake, Just, & Carpenter, 1994). In this method, participants are given a test designed to measure working memory capacity, such as Reading Span, and then divided on the basis of their scores into three groups: a high-, medium-, and low-span group. Usually, the medium-span group is omitted from further analyses, and the high- and low-span groups are compared, using ANOVAs, on another measure of interest, such as reading speed or sentence comprehension. The groups are treated as though they are independent and homogeneous, and are compared to see whether they perform differently on the secondary task (Engle *et al.*, 1992). This approach, however, has some significant flaws. First, leaving out the participants in the middle of the sample ignores a large amount of data, including information about the variability of the sample. Second, this approach may in fact lead to an overestimation of the relation between variables. Selecting only extreme groups eliminates those participants whose scores would be near the mean of the sample, leaving participants whose scores have larger deviations from the mean. As a result, the correlation coefficient is likely to be larger with extreme groups (for a discussion see McCall, 1998, pp. 168–169). Third, the choice of a cutoff point seems arbitrary if inspection of a scatterplot does not suggest any natural grouping of the data. For these reasons, the correlational approach is superior unless scatterplots show a natural grouping of the data.

SENTENCE MEMORY WITHOUT WORKING MEMORY

MacDonald and Christiansen (2002) have recently presented an alternative to the currently dominant view of the relation between working memory and language processing. They claim that there is no linguistic working memory capacity separate from linguistic representations and processes. In this view, measures of language processing and measures of linguistic working memory are simply different measures of language processing skill. Individual differences supposedly resulting from differences in working memory capacity are due to differences in skill and experience with language.

MacDonald and Christiansen point out that "the fact that subjects are tested on tasks that are called 'working memory tasks' does not entail that the construct of a working memory separate from processing is a valid one." Although this statement is true, and MacDonald and Christiansen's hypothesis has the appeal of offering an alternative to relying on differences in a poorly defined and measured working memory capacity, this view does not provide a convincing alternative that explains the existing data. First, it is not clear that it is actually an alternative: The explanation for individual differences (differences in language processing skill) translates easily into working memory models such as those of Just and Carpenter (1992) or Salthouse (1990), which view working memory capacity as the interaction of storage capacity and processing efficiency. In these models, individual differences on working memory tests could be due to differences in storage capacity, processing efficiency, or both. Thus MacDonald and Christiansen's alternative could be seen as a case of differences in a specific kind of processing efficiency (reading skill) explaining individual differences in language processing tasks and linguistic working memory tasks.

MacDonald and Christiansen's account, however, addresses only the relation between language processing and linguistic measures of working memory, such as Daneman and Carpenter's (1980) Reading Span, or the auditory analog, Listening Span, both of which involve reading and remembering the final words of sentence. The fact that these tasks require reading or listening to sentences makes it plausible to suppose that any individual differences observed may be due to differences in reading skill. But this explanation would not account for the correlations that have been observed between reading comprehension and nonlinguistic tasks of working memory (Daneman & Merikle, 1996; Engle *et al.*, 1992; Turner & Engle, 1989;). A skill-via-experience account, in which better readers do better in reading comprehension and in linguistic working memory tasks, offers no explanation for correlations among linguistic and nonlinguistic working memory tasks and not explanation for correlations between working memory, as measured by these tasks, and sentence memory.

EXPERIMENT

The experiment consisted of two parts: Part 1 tested participants' sentence memory capacity. Part 2 tested participants' short-term storage and working memory capacity.

Method: Part 1: Sentence Memory

Over headphones, participants heard sentences of different lengths, from two to five clauses. Sentences were semantically unconstrained: Any agent could plausibly perform any action in the sentence. Immediately following the sentence, participants heard a question, probing their memory for one of the clauses in the sentence. To ensure that participants paid attention to all parts of the sentence, two types of probes were presented: questions probing memory for either the subject of the clause (Agent questions) or the main verb (Action questions). For all clauses in the sentences except the final clause, there were an equal number of probes at each serial position. The final clause of the sentence was never probed.

To address the question of the unit of sentence memory, three different sentence types were assessed: Sentences with relative clauses (RC), sentential complements (SC), and a relative clause with double objects (DO) (two noun phrases [NPs] or one NP and one prepositional phrase [PP]. DO sentences with the same number of clauses were longer (containing one additional NP per clause) than RC and SC sentences. In fact, DO sentences at a given length (n) contained at least as many NPs as RC and SC sentences at length $n + 1$. Furthermore, if the NPs and verbs are grouped together as discourse referents (NPs: object referents; verbs: event referents; Heim, 1982; Kamp, 1981), then DO sentences at a given length (n) still contained at least as many discourse referents (NPs and verbs grouped together) as RC and SC sentences of length $n + 1$. Thus, if sentence memory is a function of the number of NPs or discourse referents in a sentence, then accuracy on DO sentences should be worse than the other two sentence types: DO sentences of length n should be recalled as poorly as the other two sentence types at length $n + 1$. However, if memory is a function of the number of clauses in the sentence, then there should be no difference between DO sentences and the other two sentence types.

Participants

Thirty MIT students served as participants. All participants were native English speakers. Participants were paid $8 for their participation.

Materials

Each participants heard 121 sentences of four different lengths, 2-clause, 3-clause, 4-clause, and 5-clause. Three types of sentences were used: RC, SC, and DO. The Question Type (Action versus Agent) and Probe Location (1st, 2nd, 3rd, or 4th clause) variables were balanced across the sentence types, such that equal numbers of the two question types and locations were probed for each sentence type. Order was randomized, then the same set of sentences were presented to each participant. Sample 3-clause sentences of each type are presented in (6)–(8). All clauses, except the final clause, were semantically unconstrained: Sentences were constructed using a program that randomly assigned subjects with verbs for all but the final clause in the sentence. The final clause (which was never probed) was constructed to be semantically plausible to provide some conceptual closure to the sentence. Except for the final clause, each serial position was probed equally often. For RC and DO sentences, probes consisted of two question types, to optimize comprehension and attention to the entire sentence: Agent questions (Who lectured someone?) and Action questions (What did the barber do?). Because SC sentences did not have a single NP object, only Agent questions were used.

(6) Relative Clause (RC)
The barber lectured the sailor who hit the singer who worked in the jazz club.

(7) Sentential Complement (SC)
The violinist insisted that the immigrant doubted that the chef had trained in Paris.

(8) Double Object (DO)
The psychologist showed the document to the criminal who sent a gift to the editor who was compiling an anthology.

Procedure

Sentences were recorded onto a Macintosh Quadra 640 computer using Sound Designer II software and were played back to the participant over headphones. Sentences were read naturally. Immediately following sentence presentation, participants heard a question about the sentence, which they answered aloud. Participants pressed the spacebar when they were ready for the next sentence. The experimenter marked the accuracy of the participant's response on a score sheet. Agent questions were scored 1 if they were answered correctly and 0 if they were answered incorrectly. Action questions (which were asked of only the RC and DO sentences) were scored 1 if both the verb and the direct object were correct, 0.5 if one part was correct, and 0 if both parts were incorrect or omitted. To keep the difficulty similar for the two sentence types, participants did not have to report the second object of the DO sentences. The experimental session began with six practice sentences, followed by the 121 test sentences. The session lasted approximately 30 min.

Data Analysis

For RC and DO clauses, ANOVAs were performed at each clause length to determine whether there was an effect of probe question type (Action versus Agent). (Recall that only Agent questions were asked of SC sentences, so this comparison is not relevant to the SC sentences.) There were no significant differences between question types at any clause length, so the two question types were combined for all further analyses.

Results

Figure 1 plots sentence memory as a function of the number of clauses for the three sentence types. Error bars show 95% confidence intervals. The means, standard deviations, and ranges are shown in Table I. Significant main effects were found for number of clauses, $F(3,87) = 322.99$, $p < .001$, and sentence type, $F(2,58) = 4.34$, $p < .02$; the interaction was also significant, $F(6,174) = 2.62$, $p < .05$. The difference between 4- and 5-clause sentences was significant overall, $F(1,29) = 10.5$, $p < .005$. This difference reflects the fact that accuracy on RC and DO sentences continued to decline as sentence length increased (RC: 4-clause, 36%; 5-clause, 24%; DO: 4-clause, 32%; 5-clause, 23%) whereas accuracy on SC sentences did not (4-clause, 26%; 5-clause, 27%).

One of the primary questions of interest was whether accuracy on DO sentences, in which each clause contained one additional NP (also a new

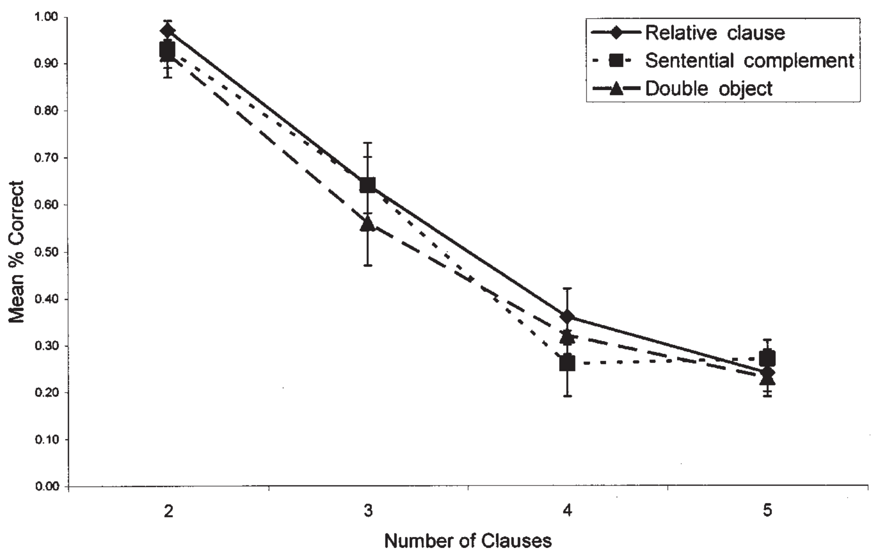

Fig. 1. Accuracy of sentence memory as a function of sentence type and number of clauses.

Table I. Percent Correct for Each Sentence Type and Length: Means, (Standard Deviations), and Ranges

	2 Clauses	3 Clauses	4 Clauses	5 Clauses
Relative Clause (RC)	97 (5) 82–100	64 (17) 31–94	36 (17) 0–79	24 (13) 3–53
Sentential Complement (SC)	93 (12) 67–100	64 (26) 17–100	26 (19) 0–67	27 (12) 8–50
Double Object (DO)	92 (13) 56–100	56 (24) 13–100	32 (12) 8–58	23 (9) 6–41

discourse referent), would be significantly worse than accuracy on the other two types. Figure 1 shows that error bars overlapped at all sentence lengths, suggesting that there was no reliable difference between sentence types. Planned comparisons showed marginally significant differences between the sentence types at the 2-clause [$F(2,58) = 2.72$, $p < .10$] and 3-clause [$F(2,58) = 2.8$, $p < .10$] lengths, but no consistent finding of significantly impaired performance on DO sentences relative to the other sentence types. Instead, accuracy declined as a function of the number of clauses for all three sentence types. These results suggest that sentence memory is sensitive to the number of clauses in a sentence, and not the number of NPs or discourse referents.

Method: Part 2: Working Memory

Several different working memory measures were used, in order to explore the relation among them and to sentence memory.

Participants

Twenty-six participants who had participated in Part 1 returned on a second day to participate in Part 2.

Materials

Short-Term Storage Tests

Forward Digit Span. Participants heard strings of digits presented at the rate of one digit per second, and then recalled them in order. Span was defined as the longest string of digits a participant could repeat correctly, in order, on one of two trials.

Word Span. Participants heard lists of words (one-syllable concrete nouns) presented at the rate of one word per second, and then recalled them in order. Span was defined as the longest string of words a participant could repeat correctly in order on one of two trials.

Working Memory Tests

Backward Digit Span. Participants heard strings of digits presented at the rate of one digit per second, and then recalled them in reverse order. Span was defined as the longest string of digits a participant correctly repeated in reverse order on at least one of two trials.

N-back. Participants saw words (four-letter abstract nouns) presented one at a time on the computer screen at the rate of one word every three seconds (2500-msec word presentation, 500-msec interstimulus interval). They responded with a button press whenever they saw a target. A target was defined as a word that was the same as the word presented N ago, or "N-back" (for discussion of this task, see Dobbs & Rule, 1989; Smith & Jonides, 1997). Participants were first presented with 2-back targets; if they reached criterion (70% correct) they were presented with 3-back targets, then 4-back, and then 5-back. There were 70–80 trials at each set size, with 10 correct targets (hits) per set. The score for each set size was computed by subtracting the number of false alarms from the number of hits to correct for guessing; then scores for all levels completed were combined to reach a composite N-back score. The equation for combination was as follows: 1 + [(2-back, % correct) + (3-back, % correct) + (4-back, % correct) + (5-back, % correct) × 100].

Reading Span. (Daneman & Carpenter, 1980). Participants viewed sets of short declarative sentences (5–10 words, mean 7.3) on the computer screen and read them aloud. Next, participants viewed simple questions (probing either the subject or the main verb) and answered them aloud. After two sentence-question sets, participants were prompted to recall the final word of both sentences. Participants were first presented with five trials at set-size two; to advance to larger set sizes (three to six), they had to recall all the words correctly on three of the five trials. Span was defined as the largest set size at which participants recalled all of the words correctly on four of the five trials; with an additional 0.2 added for each trial they recalled correctly at the next set size.

Math Span. The procedure was the same as for Reading Span, except that participants saw a simple addition problem and reported the sum aloud. After two such trials, participants recalled the second digit of each of the two problems aloud. If the participant recalled the two digits correctly on three of five trials at set size two, the set size was increased to three. The largest possible set size was six. Scoring was the same as for Reading Span.

Category Span. The procedure was the same as for Reading Span, except that participants read a list of four nouns, three of which belonged to a common category (i.e., animals, foods, or colors). The fourth word did not match the category. On each trial, participants reported the category name aloud. After two such trials, participants recalled the mismatch word for the two lists. If the participant recalled the two words correctly on three of five trials at set size two, the set size was increased to three. The largest possible set size was six. Scoring was the same as for Reading Span.

Counting Span. The procedure was the same as for Reading Span, except that participants saw sets of yellow and blue dots on the screen, counted the yellow dots, and reported the number aloud. After two such trials, participants said aloud the number of yellow dots that they had counted on each of the two screens. If the participant recalled the two numbers correctly on three of five trials at set size two, the set size was increased to three. The largest possible set size was six. Scoring was the same as for Reading Span. Notice that, unlike the three above tasks, in this task participants are asked to recall the same information (the number of yellow dots) at the end of a set as they recalled after each trial. In the previous tests, the items to recall following the set of trials was different from the item reported on each trial.

Procedure

Participants were tested in the following order for the first five memory tests: Forward Digit Span, Backward Digit Span, Word Span, N-back. Testing order was held constant for all participants because the primary measures of interest were correlations between scores, thus it was important to hold constant such factors as practice and fatigue effects. Then, the four complex span measures were presented in pseudo-random order, with the condition that the two tests requiring word recall (Reading Span and Category Span) and the two tests requiring number recall (Math Span and Counting Span) never occurred consecutively. Testing order for these for the complex span measures was randomized because scores were combined into a composite that would equally reflect the contribution of each score. Thus, fatigue and practice effects should be distributed equally across the four tests. The experimental session lasted approximately 1 hour.

Data Analysis

For the purpose of correlations between the sentence memory task and the working memory tasks, the mean score on the 3-clause sentences was used, because these scores showed the largest individual differences. Scores at the other sentence lengths might have restricted ranges because of ceiling and floor effects. Correlations for the mean overall sentence memory score and the mean score on 3-clause sentences are reported in Table II. For the tests discussed below, there is no difference in results for the 3-clause score versus the overall score.

Results

Table II shows zero-order correlations among all working memory measures, the composite Complex Span score, the overall sentence memory score, and the mean sentence memory score for 3-clause sentences.

To facilitate interpretation of the results and reduce the number of comparisons to a statistically permissible level, we focus on a subset of the working memory tests. The subset to be analyzed and discussed in depth is: Word Span, Backward Digit Span, N-back, and a Complex Span measure derived by forming a composite score from the three tests modeled on Daneman and Carpenter's (1980) Reading Span test: Reading Span, Math Span, and Category Span.

Of the short-term storage tests, we focus on Word Span rather than forward Digit Span only because it correlated better with sentence memory.

Table II. Zero-Order Correlations for All Tests

	WS	bDS	N-back	Comp span	Read span	Math span	Cat span	Count span	Sent mem (all)	Sent mem (3 Clauses)
fDS	0.29	0.71[a]	0.20	0.40[b]	0.25	0.34	0.39[b]	0.22	0.29	0.31
WS		0.44[b]	0.21	0.57[c]	0.52[b]	0.28	0.58[b]	0.12	0.43[b]	0.44[b]
BDS			0.43[b]	0.46[b]	0.32	0.44[b]	0.37	0.19	0.48[c]	0.50[c]
N-back				−0.09	−0.15	0.1	−0.17	0.24	0.49[c]	0.48[c]
Comp Span					0.88[a]	0.74[a]	0.81[a]	0.28	0.44[b]	0.48[e]
Read Span						0.48[c]	0.65[c]	0.23	0.31	0.33
Math Span							0.33	0.28	0.37	0.49[c]
Cat Span								0.17	0.39[b]	0.36
Count Span									0.16	0.31
SentMem (all)										0.92[a]

[a]$p < .001$.
[b]$p < .05$.
[c]$p < .01$.

The results of other analyses are worse when forward Digit Span is used rather than Word Span.

The composite Complex Span score was formed by first computing a z score for each participant on Reading Span, Math Span, and Category Span, then computing the linear combination of the z scores on the three tests. The linear combination of z scores was used so that each test contributed equally to the Complex Span score. Counting Span was not included in the Complex Span score because of the difference in procedure described above and because scores on Counting Span were not significantly correlated with any other test (see Table II).

The composite score was used to obtain a measure of simultaneous storage and processing capacity that was independent of the processing task. Using the composite score also increased the power of the analyses of the Span tasks, because the composite score is based on three times as much data from each participant than each of the individual span tasks. Note that the task demands of the three tests were identical. The only difference between them was the background processing task and the item to recall. Combining several tasks into a composite measure rules out the possibility that any observed relation between the variables might be due to the materials used in this study. It also makes it unlikely that the composite span task and sentence memory may both be sensitive to a third factor, a participant's reading skill (MacDonald & Christiansen, 2002), because Math and Category Spans are not reading tasks. Controlling for these alternative explanations is especially important given that there are so many versions of the original Reading Span measure, and very few attempts have been made to validate and compare different versions (Baddeley *et al.*, 1985; Daneman & Merikle, 1996; Turner & Engle, 1989; Waters & Caplan, 1996).

Table III shows correlations among the working memory measures (Backward Digit Span, N-back, and Complex Span) and an STS measure (Word Span). All correlations were significant except the correlation between N-back and Word Span, and N-back and Complex Span.

Table IV shows correlations between working memory and Word Span (an STS test) and sentence memory capacity. All correlations were significant.

Table III. Correlations Among Working Memory and Short-Term Storage Measures

	Back digit span	N-back	Complex span
Word Span	.44[a]	.21	.57[b]
Back digit Span		.43[a]	.46[a]
N-back			−.09

[a]$p < .05$.
[b]$p < .01$.

Table IV. Correlations Among Sentence Memory Capacity and Working Memory and Short-Term Store Measures

STS/WM test	Correlation with sentence memory
Word Span	.44[a]
Back digit Span	.50[b]
N-back	.48[b]
Complex Span	.48[b]

[a] $p < .05$.
[b] $p < .01$.

To determine the relative contribution of these variables to explaining individual differences in working memory capacity, the four variables were entered into a multiple stepwise regression equation (Table V): The linear combination of Backward Digit Span, Word Span, N-back, and Complex Span significantly predicted sentence memory [$F(4,21) = 5.57$, $p < .005$], explaining 51% of the variance in the sentence memory score (Multiple R^2). However, the only two predictors that contributed to the relation were N-back, which uniquely accounted for 16% of the variance, and Complex Span, which accounted for an additional 12% (squared semi-partial correlations). In spite of correlating significantly with sentence memory, Backward Digit Span and Word Span did not account for any additional unique variance. When N-back and Complex Span were entered alone as predictors, the model still accounted for 51% of the variance in sentence memory, and each predictor explained 28% of the unique variance in sentence memory capacity (Table VI).

DISCUSSION

The first topic addressed in this paper was the unit of sentence memory. Two alternative hypotheses were tested: (i) that sentence memory would be

Table V. Multiple Regression of Working Memory and STS Variables: Percentage of Variance in Sentence Memory and Significant *p* Values

	R^2 for model = .51	
STS/WM test	% Variance[a]	*p*
N-back	16	.01
Complex Span	12	.03
Word Span	<1	n.s.
bDS	<1	n.s.

[a] Squared semi-partial correlations.
n.s., not significant

Table VI. Multiple Regression of N-back and Complex Span: Percentage of Variance in Sentence Memory and Significant *p* Values

	R^2 for model = .51	
STS/WM Test	% Variance[a]	*p*
N-back	28	.001
Complex Span	28	.001

[a]Squared semi-partial correlations.

a function of the number of clauses; or (ii) that sentence memory would be a function of the number of words, NPs, or new discourse referents. The results of the experiment suggest that participants recalled the content of the sentences as a function of the recency of presentation of the number of clauses in a sentence, and not the recency of presentation of the number NPs or discourse referents in a sentence. Thus, although new discourse structure appears to be an important measure of locality in on-line sentence comprehension (Gibson, 1998), the current results suggest that the clause is a more important storage unit for sentence memory. This finding thus confirms the hypothesis from the literature that the unit of sentence memory is the clause (Blauberg & Braine, 1974; Jarvella, 1971). Furthermore, this finding is relevant to the debate about whether on-line language processing and off-line sentence memory use the same memory resources as proposed by Just and Carpenter (1992) or different pools of resources, as proposed in Caplan and Waters (1999). The fact on-line processing and sentence memory use different representational units is consistent with Caplan and Water's suggestion that the two types of processing are distinct.

The second question addressed in this paper concerned the relation between sentence memory and working memory. Half of the variance in sentence memory capacity could be explained by a combination of two working memory tests, the N-back and the Complex Span measure, a composite of three tests modeled on Daneman and Carpenter's (1980) original Reading Span measure. Because Complex Span is a combination of three tests with similar task demands (simultaneously storing and processing information and switching attention between the two subtasks) but with different processing requirements (reading sentences, performing addition problems, categorizing words), the predictive power of Complex Span reflects the relation between sentence memory and whatever aspect of working memory that Complex Span measures. Because neither Complex Span nor N-back are tasks that involve reading sentences (with the exception of the Reading Span component of Complex Span), this finding casts doubt on MacDonald and Christiansen's (2002) hypothesis that correlations between linguistic working memory measures and sentence comprehension measures are due to the fact that both are sensitive to participant's reading ability.

These results are also relevant to MacDonald and Christiansen's attempt to abolish the working memory construct. Their alternative skill-via-experience account, such that better readers are better at both linguistic working memory and sentence comprehension tasks, does not account for the correlations observed in this study. N-back, Math Span, and Category Span, tests that did not involve reading sentences, were correlated with sentence memory. In fact, the zero-order correlations (shown in Table II) between Math Span and Category Span and sentence memory were higher than between Reading Span and sentence memory. MacDonald and Christiansen's explanation for such correlations is that all these tests are sensitive to "the accuracy of phonological representations," which they claim is a biological factor underlying individual differences (along with reading skill, an experiential factor). In fact, their explanation for individual differences on Reading Span and Listening Span is that these tests reflect not only differences in reading ability, but also differences in phonological processing ability, rather than differences in working memory. They state that "maintaining a set of unrelated words requires substantial activation of phonological representations" (p. 45). Maintaining phonological activation of words, however, is another way of describing the storage functions of working memory. Thus, MacDonald and Christiansen have not, in fact, presented an adequate alternative to the idea of variance in working memory capacity as the source of individual differences in understanding language, but have restated the problem. The current finding, that there is a correlation between working memory tests and sentence memory, suggests that they are both calling on resources that are central, in that the same resources are used for a variety of complex cognitive tasks.

REFERENCES

Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppe, R. A., & Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory: Evidence from Positron Emission Tomography. *Psychological Science, 7,* 25–31.

Babcock, R. L., & Salthouse, T. A. (1990). Effects of increased processing demands on age differences in working memory. *Psychology and Aging, 5,* 421–428.

Baddeley, A. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology, 49A,* 5–28.

Baddeley, A. D. (1983). Working Memory. *Philosophical Transactions of the Royal Society of London, B 302,* 311–324.

Baddeley, A. D., Bressi, S., Sala, R. D., Logie, R., & Spinnler, H. (1991). The decline of working memory in Alzheimer's disease: A longitudinal study. *Brain, 114,* 2521–2542.

Baddeley, A., Eldridge, M., & Lewis, V. (1981). The role of subvocalisation in reading. *Quarterly Journal of Experimental Psychology, 33A,* 439–454.

Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology, 8,* 485–493.

Baddeley, A. D., Logie, R., Bressi, S., Sala, R. D., & Spinnler, H. (1986). Dementia and working memory. *Quarterly Journal of Experimental Psychology, 38A,* 603–618.

Baddeley, A., Logie, R., Nimmo-Smith, I., & Brereton, N. (1985). Components of fluid reading. *Journal of Memory and Language, 24,* 119–131.

Baddeley, A., Vallar, G., & Wilson, B. (1987). Sentence comprehension and phonological memory: Some neuropsychological evidence. In M. Coltheart (Ed.), *Attention and Performance XII.* (509–529). Hove, UK: Lawrence Erlbaum.

Basso, A., Spinnler, H., Vallar, G., & Zanobio, M. E. (1982). Left hemisphere damage and selective impairment of auditory verbal short-term memory: A case study. *Neuropsychologia, 20,* 263–274.

Belleville, S., Peretz, I., & Arguin, M. (1992). Contribution of articulatory rehearsal to short-term memory: Evidence from a case of selective disruption. *Brain and Language, 43,* 713–746.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the Development of language* (279–362). New York: Wiley.

Bever, T. G., Lackner, J. R. & Kirk, R. (1969). The underlying structures of sentences are the primary units of immediate speech processing. *Perception & Psychophysics, 5,* 225–234.

Blauberg, M. S., & Braine, M. D. S. (1974). Short-term memory limitations on decoding self-embedded sentences. *Journal of Experimental Psychology, 102,* 745–748.

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage, 5,* 49–62.

Caplan, D. (1972). Clause boundaries and recognition latencies for words in sentences. *Perception & Psychophysics, 12,* 73–76.

Caplan, D., & Waters, G. S. (1990). Short-term memory and language comprehension: A critical review of the neuropsychological literature. In G. Vallar & T. Shallice (Eds.), *Neuropsychological Impairments of Short-term Memory* (337–389). Cambridge: Cambridge University Press.

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

Case, R., Kurland, E. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33,* 386–404.

Chang, F. R. (1980). Active memory processes in visual sentence comprhension: Clause effects and pronominal reference. *Memory & Cognition, 8,* 58–64.

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

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating informatin between and within sentences. *Journal of Experimental Psychology: Learning Memory and Cognition, 9,* 561–584.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3,* 422–433.

D'Esposito, M., Detre, M. A., Alsop, D. C., Shin R. K., Atlas, S., & Grossman, M. (1995). The neural basis of the central executive system of working memory. *Nature, 378,* 279–281.

Dobbs, A. R., & Rule, B. G. (1989). Adult age differences in working memory. *Psychology and Aging, 4,* 500–503.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory & Cognition, 18,* 972–992.

Fiez, J. A., Raife, E. A., Balota, D. A., Schwartz, J. P., Raichle, M. E., & Petersen, S. E. (1996). A positron emission tomography study of the short-term maintenance of verbal information. *The Journal of Neuroscience, 16,* 808–822.

Frazier, L. (1978). On comprehending sentences: Syntactic parsing strategies, Ph.D. thesis, University of Connecticut, Storrs, CT.
Frazier, L., & Rayner, K. (1988). Parameterizing the language-processing system: Left-vs. right-branching within and across languages. In J. A. Hawkins (Ed.), *Explaining language universals* (247–279). Oxford, UK: Blackwell.
Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language.* Hillsdale: Lawrence Erlbaum.
Gershberg, F. B., & Shimamura, A. P. (1995). Impaired use of organizational strategies in free recall following frontal lobe damage. *Neuropsychologia, 13,* 1305–1333.
Gibson, E. (1991). A computational theory of human linguistic processing: Memory limitations and processing breakdown. Unpublished doctoral dissertation, Carnegie-Mellon University, Pittsburgh, PA.
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 Miyashita, Y., Marantz, A., & O'Neil, W. (Eds.), *Image, language, brain* (95–126), Cambridge, MA: MIT Press.
Glanzer, M., & Razel, M. (1974). The size of the unit in short-term storage. *Journal of Verbal Learning and Verbal Behavior, 13,* 114–131.
Heim, I. (1982). The semantics of definite and indefinite noun phrases. PhD Thesis. University of Massachusetts, Amherst, MA.
Jarvella, R. J. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior, 10,* 409–416.
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., & Keller, T. A. (1996). The capacity theory of comprehension: New frontiers of evidence and arguments. *Psychological Review, 103,* 773–780.
Kac, M. B. (1981). Center-embeddings revisited. *Proceedings of the third annual conference of the Cognitive Science Society.* Hillsdale: Lawrence Erlbaum.
Kamp, H. (1981). A theory of truth and semantic representations. In J. Groenendijk (Ed.) *Formal methods in the study of language.* Mathematisch Centrum, Amsterdam, The Netherlands.
Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition, 2,* 15–47.
King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30,* 580–602.
King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7,* 376–395.
Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85,* 363–394.
Klapp, S. T., Marshburn, E. A., & Lester, P. T. (1983). Short-term memory does not involve the "working memory" of information processing: The demise of a common assumption. *Journal of Experimental Psychology: General, 112,* 240–264.
LaPointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16,* 1118–1133.
Larkin, W., & Burns, D. (1977). Sentence comprehension and memory for embedded structure. *Memory & Cognition, 5,* 17–22.
Lehto, J. (1996). Are executive function tests dependent on working memory capacity. *The Quarterly Journal of Experimental Psychology, 49A,* 29–50.

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

Lombardi, L., & Potter, M. C. (1992). The regeneration of syntax in short-term memory. *Journal of Memory and Language, 31,* 713–733.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: A comment on Just & Carpenter (1992) and Waters & Caplan (1996). *Psychological Review, 109,* 35–54.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology, 26,* 56–98.

Martin, R. C. (1987). Articulatory and phonological deficits in short-term memory and their relation to syntactic processing. *Brain and Language, 32,* 159–192.

Martin, R. C. (1993). Short-term memory and sentence processing: Evidence from neuropsychology. *Memory & Cognition, 21,* 176–183.

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

McCall, (1998). *Fundamental statistics for the behavioral sciences*, 7th ed. New York: Brooks/Cole.

McCarthy, R. A., & Warrington, E. K. (1987a). The double dissociation of short-term memory for lists and sentences. *Brain, 110,* 1545–1563.

McCarthy, R. A., & Warrington, E. K. (1987b). Understanding: A function of short-term memory? *Brain, 110,* 1565–1578.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

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

Miyake, A., Carpenter, P. A., & Just, M. A. (1994). A capacity approach to syntactic comprehension disorders: Making normal adults perform like aphasic patients. *Cognitive Neuropsychology, 11,* 671–717.

Miyake, A., Carpenter, P. A., & Just, M. A. (1995). Reduced resources and specific impairments in normal and aphasic sentence comprehension. *Cognitive Neuropsychology, 12,* 651–679.

Miyake, A. K., Just, M. A., & Carpenter, P. A. (1994). Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple interpretations in neurtral contexts. *Journal of Memory and Language, 33,* 175–202.

Paulesu, E., Frith, C. D., & Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature, 362,* 342–345.

Petrides, M., Alivisatos, B., Meyer, E., & Evans, A. C. (1993). Functional activation of the human frontal cortex during the performance of verbal working memory tasks. *Proceedings of the National Academey of Sciences, 90,* 878–882.

Potter, M. C. (1993). Very short-term conceptual memory. *Memory & Cognition, 21,* 156–161.

Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language, 29,* 633–654.

Potter, M. C., Moryadas, A., Abrams, I., & Noel, A. (1993). Word perception and misperception in context. *Journal of Experimental Psychology: Learning, Memory and Cognition, 19,* 3–22.

Potter, M. C., Valian, V. V., & Faulconer, B. A. (1977). Representation of a sentence and its pragmatic implications: Verbal, imagistic, or abstract? *Journal of Verbal Learning and Verbal Behavior, 16,* 1–12.

Roberts, R. M., & Corkin, S. (1997). Poor correlation among verbal working memory tests. Poster presented at the Cognitive Neuroscience Society Meeting, Boston, MA.

Saffran, E. M., & Marin, O. S. M. (1975). Immediate memory for word lists and sentences in a patient with deficient auditory short-term memory. *Brain and Language, 2,* 420–433.

Salthouse, T. A. (1990). Working memory as a processing resource in cognitive aging. *Developmental Review, 10,* 101–124.

Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology, 33,* 5–42.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. L. Miller & P. D. Eimas, (Eds.), *Speech, language and communication* (217–262). San Diego, CA: Academic Press.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28,* 127–154.

Vallar, G., Betta, A. M. D., & Silveri, M. C. (1997). The phonological short-term store-rehearsal system: Patterns of impairment and neural correlates. *Neuropsychologia, 35,* 795–813.

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

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology, 49A,* 51–79.

Waters, G. S., & Caplan, D. (1997). Working memory and on-line sentence comprehension in patients with Alzheimer's disease. *Journal of Psycholinguistic Research, 26,* 377–400.

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