Journal of Psycholinguistic Research, Vol. 29, No. 2, 2000

Sentence Comprehension Is Mediated by Content-Addressable Memory Structures

Brian McElree[1]

Studies of working memory demonstrate that some forms of information are retrieved by a content-addressable mechanism (McElree & Dosher, 1989; McElree, 1996, 1998), whereas others require a slower search-based mechanism (McElree & Dosher, 1993). Measures of the speed and accuracy of processing sentences with filler-gap dependencies demonstrate that the probability of maintaining a representation of a filler item decreases as additional material is processed, but that the speed with which a preserved representation is accessed is unaffected by the amount of interpolated material. These results suggest that basic binding operations in sentence comprehension are mediated by a content-addressable memory system.

Language comprehension requires constructing a hierarchical representation from a linear sequence of symbols. This mapping entails resolving dependencies between constituents that can be separated by an indefinite amount of material. To resolve nonadjacent dependencies, on-line processes must have access to constituents processed at earlier points. Working memory (WM) is thought to provide the "work space" where the products of prior analyses can be maintained in an accessible and modifiable format.

The study reported here examines how previously processed constituents are accessed during on-line comprehension as a means of exploring the nature of the WM representations that subserve language processing. The data suggest that basic binding operations in comprehension are mediated by memory representations that are content addressable: syntactic and semantic constraints provide direct access to relevant representations without the need to search through potentially irrelevant information. Such a system

This research was supported by NIMH Grant MH57458. The author thanks Lisbeth Dyer and Stephani Foraker for assistance with the generation of materials and data collection.

[1] Program in Cognition and Perception, Department of Psychology, New York University, 6 Washington Place, New York, 10003-6634. Email: bdm@psych.nyu.edu.

0090-6905/00/0300-0111$18.00/0

contrasts with traditional views of WM and with recent findings that access to some forms of information in WM requires a search process.

ACCESSING WM STRUCTURES

The behavioral marker of a search process is that the time to access a representation depends on the amount of information held in memory or the amount of information interpolated between initial processing and retrieval. Myriad studies demonstrate that the probability of maintaining a representation in WM decreases as a function of both factors. However, a search process crucially entails that search time will also vary over and above differences in the probability of maintaining a representation. I'll refer to the probability that a memory representation is maintained as the *availability* of the representation and the time it takes to retrieve the representation (conditional on its being available) as the *accessibility* of the representation.

In a content-addressable mechanism, in contrast, cues at retrieval provide direct access to relevant memory representations without the need to search through extraneous information. Although representations may differ in availability, a content-addressable system enables representations of differing quality or strength to be retrieved in equal time. Content-addressable retrieval operations can be implemented in memory models with rather diverse storage architectures, including models with highly structured localized representations and those with highly distributed representations (for a review, see Clark & Gronlund, 1996).

Given the capacity of long-term memory, most viable models posit content-addressable retrieval processes (Gillund & Shiffrin, 1984; Hintzman, 1984; Murdock, 1982; see also Clark & Gronlund, 1996). However, a search process cannot be precluded on a priori grounds in memory systems with small capacity. Traditional models of WM propose that retrieval does indeed involve a search operation, often of a serial form (Sternberg, 1975; see also Dosher & McElree, 1992; McElree & Dosher, 1993). That position has been modified by recent studies demonstrating that some forms of information are retrieved with a serial and others with a content-addressable operation.

McElree and Dosher (1989; see also McElree, 1996, 1998; McElree & Dosher, 1993; Wickelgren *et al*., 1980) examined retrieval operations in short-term item recognition, the paradigmatic case for studying how representations in WM are accessed. Using a speed–accuracy tradeoff (SAT; see below) procedure, availability was found to be adversely affected by the number of items in memory and the number of items interpolated between study and test, but accessibility was unaffected by both factors. Search models, whether serial or parallel in form, are incompatible with findings that representations of differing availability are nevertheless equally accessible. These data indicate that WM representations are accessed by a content-addressable operation.

A content-addressable operation does not, however, enable the recovery of all forms of information. Subsequent work demonstrated that relational information, including temporal order (McElree & Dosher, 1993) and positional (Gronlund *et al.*, 1997) information, requires a (slower) serial search. For example, in a study of the retrieval of temporal order information, McElree and Dosher (1993) documented that both availability and accessibility depend on the amount of information interpolated between study and test. Retrieving relational information requires a systematic search over an ordered set of representations.

WM STRUCTURES IN SENTENCE COMPREHENSION

Is access to a previously processed constituent in on-line comprehension mediated by a content-addressable mechanism or by a systematic search through prior representations? This issue was addressed by contrasting the speed and accuracy of processing filler-gap constructions of the type illustrated in 1–3.

1. This was the book that the editor admired. (*amused.)
2. This was the book that the editor who the receptionist married admired. (*amused.)
3. This was the book that the editor who the receptionist who quit married admired, (*amused.)

The fronted NP (*the book*) in each case must be bound to the gap in the direct object position of the final matrix verb (*admired*) to correctly interpret the strings. The strings differ in the amount of material that must be processed prior to the gap in constituent structure. The filler and gap in (1) are separated by the matrix subject (*the editor*) only. In (2), an object relative clause (*who the receptionist married*) attached to the matrix subject is interpolated between the filler and gap. In (3), an additional subject relative clause (*who quit*) is attached to the subject of the prior relative clause, further increasing material between the filler and gap. Construction (3) is close to the upper-bound on processing embedded structures.

If binding of the filler item to the matrix verb requires a search—either through a representation of the linear surface structure or through a more structured representation—then processing speed should systematically slow as more material intervenes between the filler and gap. Such a pattern might be expected if binding operations crucially depend on relational information (e.g., McElree & Dosher, 1993). Alternatively, processing speed should not differ across the strings if memory structures are content addressable *and* syntactic and semantic constraints available to the parser can be used to access those memory structures.

Measures of Processing Speed

To measure the speed of binding a filler item, it is necessary to use a task that requires full interpretation of the strings. The strategy here, as in other work (McElree & Griffith, 1995, 1998), was to use a task that required subjects to discriminate acceptable from unacceptable bindings. Unacceptable versions of (1–3) used final matrix verbs (**amused*) that required an animate filler NP. Subjects were required to make binary ("yes/no") acceptability decisions at varying times after processing the crucial matrix verb.

The representation for the filler item was expected to be adversely affected by the amount of interpolated material, based on standard findings in memory research and particular models of memory for sentence processing (e.g., Lewis, 1996). Discriminating between content-addressable and search processes requires measures of processing speed that are unaffected by differences in memory availability. (Simple timing measures like reaction or reading time are affected by both factors; see McElree, 1993; McElree & Nordlie, 1999; McElree & Griffith, 1995, 1998). The standard solution to this problem is to derive a function that measures how accuracy varies with processing time (Wickelgren, 1977). Here, the response-signal SAT procedure was used to construct such functions.

Figure 1 presents an overview of the SAT procedure. Sentences were visually presented one word at time. Acceptability judgments were required at one of six times that varied (between trials) from 50 to 3000 ms after the onset of the final word. When to respond was signaled by a brief tone and subjects were trained to respond within 300 ms of the tone. In the main contrasts of interest (1–3), the final word before the response tone was the matrix verb, which projected an argument position for the fronted NP. SAT functions were derived by measuring accuracy (d' units) as a function of processing time (time of the tone plus response latency). SAT functions provide a measure of when accuracy departs from chance (an intercept), the rate at which accuracy grows over processing time and the maximum or asymptotic level of performance (see illustrative curve in lower right panel in Fig. 1).

The level of asymptotic performance is determined by the probability that a representation of the filler was available and that material interpolated between the filler and gap was successfully parsed. The crucial measure, speed of processing, is jointly assessed by the point at which accuracy departs from chance (intercept) and the rate at which accuracy grows to asymptote.

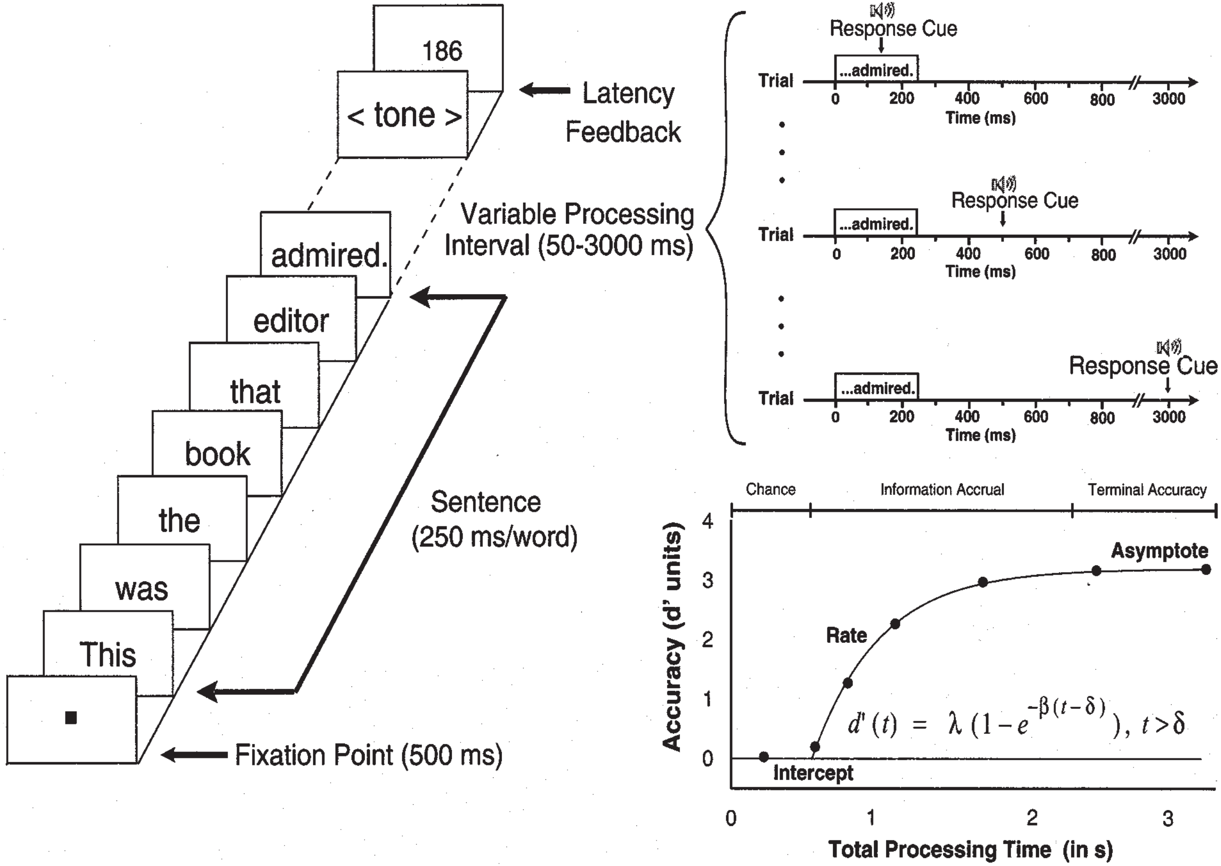

Fig. 1. Upper left and right panels: Sample trial sequence illustrating the speed-accuracy tradeoff (SAT) variant of the acceptability judgment task. Bottom right panel: hypothetical SAT curve.

If access to the filler's representation requires a search process when the matrix verb is encountered, then the intercepts and/or rates of SAT functions will systematically slow as more material is interpolated between the filler and gap. For example, McElree and Dosher (1993) found that a search process engendered differences in SAT intercepts that were as large as 500 ms in short lists of six words. In contrast, equal intercepts and rates are indicative of a content-addressable operation.

METHOD

Subjects

SAT studies are designed to collect stable functions for individual subjects, enabling conclusions to be drawn for each subject. To that end, six students from New York University served as subjects. Each subject participated in ten 1-h sessions, plus a 1-h practice session.

Materials

Ten sets of 576 sentences composed of 24 distinct sentence types were used.[2] The main contrasts are given in (1–3). The sets consisted of an equal number of acceptable and unacceptable sentences and an equal number of sentences without displaced constituents but with otherwise comparable complexity to the filled-gap constructions (e.g., _The editor admired the book; The editor who the receptionist married admired the book; The editor who the receptionist who quit married admired the book_). Several constructions were included to prevent subjects from adopting specialized response strategies. For example, acceptable and unacceptable filler-gap constructions with animate initial NPs were used to prevent subjects from associating a "yes" or "no" response to particular verbs (e.g., _It was the editor that the book amused; It was the editor that the book *admired_). These constructions were equal in number and length to the main experimental constructions illustrated in (1–3). In addition, the violations in unacceptable strings involved constituents other than the fronted NP and final gap position, so that subjects could not focus exclusively on that relationship (e.g., _The_ book admired* _the editor; The_ book _which the journalist who died wrote_ admired* _the editor; It was the editor that the_ book _which the journalist wrote_ admired* _The editor who the receptionist married_ amused _the_ book*).

[2] The full set of materials is available from the author.

Procedure

Stimulus presentation, timing, and response collection were all carried out on a personal computer using software with millisecond timing. A trial began with a 500-ms fixation point (a small filled square) presented in the center of the screen. Words were presented one after another for 250 ms, with a period appended to the final word of a string. A 50-ms, 1000-Hz tone sounded at one of six response lags, either 50, 300, 500, 800, 1200, or 3000 ms after the onset of the final word in the string. Participants were trained to respond "yes" or "no" at the tone by pressing one of two designated keys on the keyboard. After a response, visual feedback on the latency to respond to the tone was given. The participants were informed that responses longer than 300 ms were unacceptably long and that responses shorter than 100 ms should be regarded as anticipations. Both the sentences and the response lags were randomized within a session.

Data Analysis

All analyses were performed on the individual subjects' data. Consistent patterns across subjects were summarized by analyses of the average data. To correct for response bias, d' scores were computed by scaling the z score of the probability of saying "yes" to acceptable strings against the z score of the probability of saying "yes" to corresponding unacceptable strings at each lag. Potential differences in asymptote, rate, and intercept were assessed by fitting the d' accuracies at various processing times (t) with an exponential approach to a limit:

$$d'(t) = \lambda\left(1 - e^{-\beta(t-\delta)}\right), \quad \text{for } t > \delta, \text{ else } 0, \tag{1}$$

where λ reflects the asymptote of the function, δ denotes the intercept or discrete point in time when accuracy departs from chance, and β indexes the rate at which accuracy grows to asymptote. Hierarchically nested models were fit to the data, ranging from a null model, in which all three conditions were fit with a single asymptote (λ), rate (β), and intercept (δ), to a fully saturated (nine parameter) model, in which each experimental condition was fit with a unique set of parameters. Fit quality was assessed by an adjusted-R^2 statistic—the proportion of variance accounted for by the fit adjusted by the number of free parameters (Judd & McClelland, 1989)—and by an evaluation of the consistency of the parameter estimates across the subjects.

RESULTS

Figure 2 shows the average (over subjects) d' data as a function of processing time for constructions in which no clause intervened between the

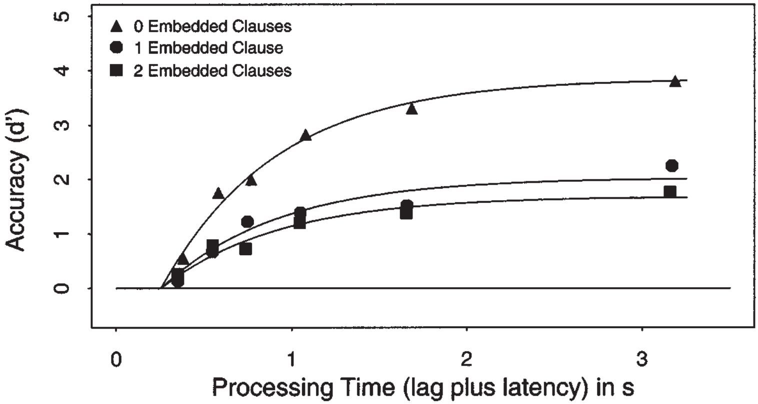

Fig. 2. Average d' accuracy (symbols) as a function of processing time (lag of the response cue plus latency to respond to the cue) for judgments of filler-gap constructions with no embedded clauses (triangles), one embedded relative clause (circles), and two embedded relative clauses (squares). Smooth curves show the best fitting 3λ–1 β–1δ exponential model (see text).

filler and gap (triangles), one embedded clause intervened between the filler and gap (circles), and two embedded clauses intervened between the filler and gap (squares). Asymptotic accuracy decreased as more material was interpolated between the filler and gap. That conclusion was supported by an ANOVA on the d' score at the longest response time (3000 ms), $F(2,10) = 39.4$, $MSE = 0.174$, and by competitive fits of Eq. (1.) A 1λ–1β–1δ (null) model, in which all three conditions were fit with a common set of parameters, produced extremely low adjusted-R^2 values, .504 in the average data ranging from .244 to .587 across the six subjects. In contrast, a 3λ–1β–1δ model, with separate asymptotes (λ) for each construction type, produced adjusted-R^2 values of .973 in the average data ranging from .756 to .946 across the six subjects. Equally important, this model yielded a consistent set of asymptotic estimates. The λ estimates for the average data were 3.85, 2.03, and 1.69 for conditions with 0, 1, 2 embedded clauses (respectively). All subjects showed estimates with this ordering, and the difference in parameter estimates was significant, $F(2,10) = 49.9$, $MSE = 0.154$. Below it is argued that these asymptotic differences partly reflect loss of the filler item because of the interfering effect of processing the embedded clauses.

Despite systematic asymptotic differences, the strings did not differ in processing speed. Two facets of the analysis compel that conclusion. First, all adjusted-R^2 values were *lower* than the 3λ–1β–1δ model when the SAT functions were fit with models that further varied the rate parameter (a 3λ–3β–1δ model), the intercept parameter (a 3λ–1β–3δ model), or both parameters (a fully saturated 3λ–3β–3δ model). Second, these models did not yield a consistent ordering of β and δ estimates. Both facts indicate that the additional parameters were not accounting for systematic variance in the data. The time-course profiles indicate that interpolated material does not affect the speed with which the antecedent filler item can be accessed and processed.

DISCUSSION

Measures of the speed and accuracy of resolving filler-gap dependencies demonstrate that the likelihood of binding a filler to the gap decreases with more interpolated material but the time to resolve the dependency, when it can be resolved, remains constant. If a search process were used to access the filler item, then processing time should have increased with more structure between the filler and gap (McElree & Dosher, 1993). To the contrary, the time-course profiles show the same pattern as simple item recognition in that availability, but not accessibility, is affected by interpolated material (McElree & Dosher, 1989; McElree, 1996, 1998; Wickelgren *et al.*, 1980). These data suggest that the representations constructed during on-line comprehension are content-addressable and that parsing operations can use syntactic and semantic constraints (see McElree & Griffith, 1998) to directly access those representations during binding operations.

The contrasts in (1–3) varied in the number of words and the complexity of the structure intervening between the filler and gap. However, vertical distance in a hierarchical (syntactic) representation is constant across (1–3) and it is possible that a search process operates over such a representation. McElree and Griffith (1998) compared strings like 4 and 5, which do differ in vertical distance:[3]

4. The writer knew this was *the essay* that the editor had admired.
5. This was *the essay* that the writer knew the editor had admired.

Figure 3 shows the d' data and best fitting exponential model for the two conditions. Although distance had a (small) effect on asymptotic accuracy, measures of accessibility were unaffected by distance. More work is needed to explore potential effects of hierarchical structure. However, these data are not indicative of a search that operates over hierarchical structure, but rather suggest that binding of the filler is mediated by a content-addressable operation.

A content-addressable notion assumes that syntactic and semantic properties available at the gap are used to *recover* a WM representation of the filler item. Alternatively, processing speed may be invariant across the strings

[3] The contrasts in (4) and (5) were noted in McElree & Griffith (1998, p. 446), but were not analyzed and reported.

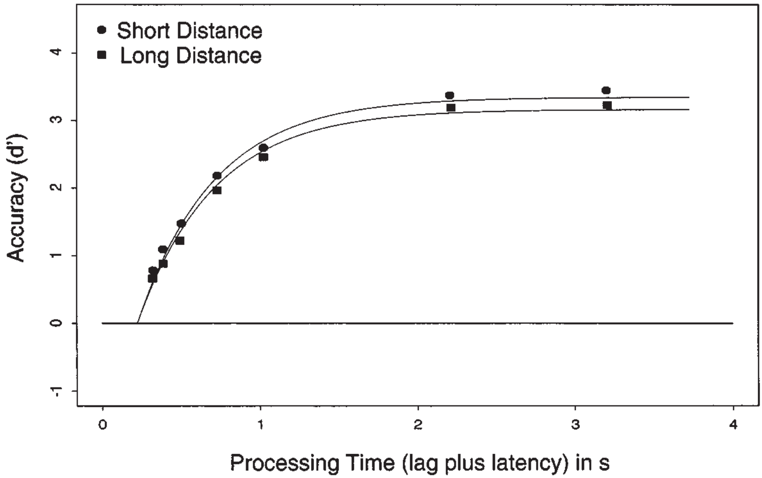

Fig. 3. Average *d′* accuracy (symbols) as a function of processing time (lag of the response cue plus latency to respond to the cue) for judgments of constructions with a short (circles) or a long (squares) distance between the filler and gap. Smooth curves show the best fitting 3λ–1 β–1δ exponential model. (Data are from unreported conditions in Experiment 3 of McElree & Griffith, 1998.)

because the parser actively maintains a representation of the filler item and so no restoration process is needed. For example, the parser may use a grammar that effectively percolates the filler through the parse tree [e.g., *slash feature* in GPSG (Gazdar *et al.*, 1985) and HPSG (Pollard & Sag, 1994)]. Such an account would have to assume that the observed asymptotic differences result from factors other than loss of availability; perhaps, for example, a greater tendency to misanalyze strings of greater complexity.

Two facts suggest that the differences observed here largely reflect availability. First, there were no significant asymptotic differences across analogs of (1–3) without a fronted direct object,[4] suggesting that complexity *per se* does not engender large differences in SAT asymptote. Second, other studies have suggested that gaps reactivate their antecedents. Cross-modal priming studies (e.g., Swinney *et al.*, 1988) have found evidence for activation of a filler after but not before the gap position, implicating reactivation of the filler by the gap. Analogously, Fig. 4 shows the asymptotic levels for a probe recognition task that we conducted in which subjects were interrupted at various points (1–5) during the reading of sentences and

[4] These contrasts included (1) *The editor admired the book;* (2) *The editor who the receptionist married admired the book;* (3) *The editor who the receptionist who quit married admired the book;* which yielded (average) asymptotic *d′*s of 3.88, 3.91, 3.67, respectively.

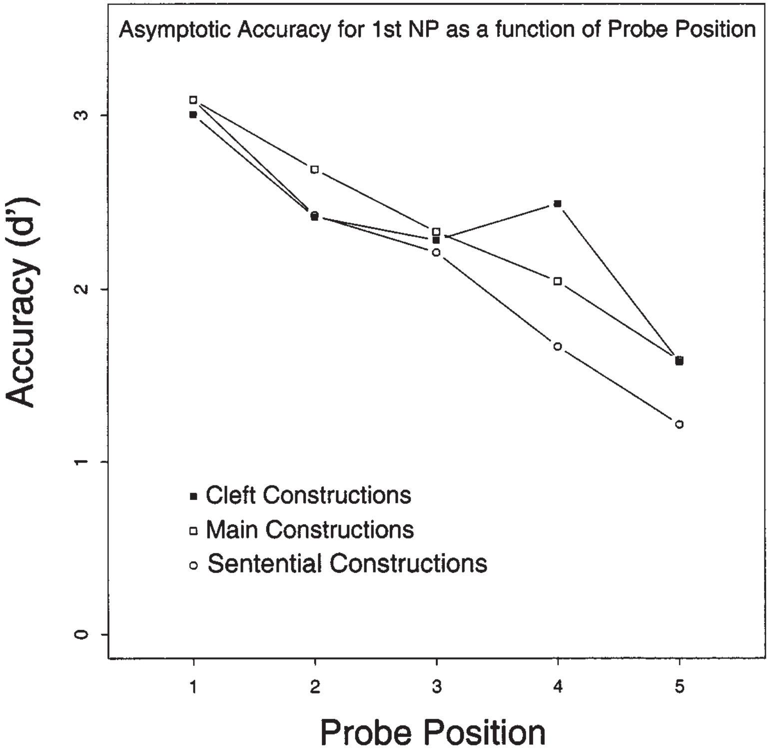

Fig. 4. Asymptotic accuracy for the (synonym) probe recognition task with cleft constructions (filled squares), simple main clause constructions (open squares), and sentential complement constructions (open circles).

required to judge whether a test probe was synonymous with an element in the sentence. Shown are the accuracies for a probe related to the initial NP in cleft constructions (*It was the fearless passengers [1] who the able sailor [2] advised [3] about the lifeboats [4] although the heavy storm was quickly abating [5]*), nonclefted main clause constructions (*The able sailor [1] had advised [2] the fearless passengers [3] about the lifeboats [4] although the heavy storm was quickly abating [5]*), and constructions with a sentential complement (*The able sailor [1] believed that [2] the fearless passengers [3] entered the lifeboats [4] although the heavy storm was quickly abating [5]*). Accuracy monotonically declined in all cases except for the cleft construction where there was a significant increase in accuracy after the gap (between positions 3 and 4). These data indicate that the representation of the initial NP is liable to interference even when it is a filler item, and that, consistent with the cross-modal priming studies, the gap reactivates a representation of the filler phrase.

The most consistent account of the cross-modal priming and probe recognition studies is one that argues for reactivation. The time-course profiles reported here indicate that reactivation is mediated by a content-addressable operation. The same time-course patterns are found in other types of binding relations that do not involve explicit filler-gap dependencies, including the binding of a subject to its matrix verb (McElree, Foraker, & Dyer, in preparation). Content addressability appears to be a general property of the WM memory structures that subserve language comprehension.

REFERENCES

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

Dosher, B. A. and McElree, B. (1992) Memory search: Retrieval processes in short-term and long-term recognition. In Larry Squire (Ed.), *Encyclopedia of learning and memory* (pp. 398–406). New York, Macmillian.

Gazdar, G., Klein, E., Pullum, G. K., & Sag, I. A. (1985). *Generalized phrase structure grammar.* Oxford: Basil Blackwell.

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

Gronlund, S. D., Edwards, M. B., and Ohrt, D. D. (1997). Comparison of the retrieval of item versus spatial position information. *Journal of Experimental Psychology: Learning, Memory, & Cognition. 23,* 1261–1274.

Hintzman, D. L. (1984). MINERVA2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers, 16,* 96–101.

Judd, C. M., & McClelland, G. H. (1989). *Data analysis: A model-comparison approach.* San Diego, CA: Harcourt Brace Jovanovich.

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

McElree, B. (1993). The locus of lexical preference effects in sentence comprehension: A time-course analysis. *Journal of Memory and Language, 32,* 536–571.

McElree, B. (1996). Accessing short-term memory with semantic and phonological information: A time-course analysis. *Memory & Cognition, 24,* 173–187.

McElree, B. (1998). Attended and non-attended states in working memory: Accessing categorized structures. *Journal of Memory & Language, 38,* 225–252.

McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: Time course of recognition. *Journal of Experimental Psychology: General, 118,* 346–373.

McElree, B., & Dosher, B. A. (1993). Serial retrieval processes in the recovery of order information. *Journal of Experimental Psychology: General, 122,* 291–315.

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

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

McElree, B., & Nordlie, J. (in press). Literal and figurative interpretations are computed in equal time. *Psychonomic Bulletin & Review, 6,* 486–494.

McElree, B., Foraker, S., & Dyer, L. (in preparation) Content-addressable memories subserve language processing.

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

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

Sternberg, S. (1975). Memory-scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology, 27,* 1–32.

Swinney, D., Ford, M., Bresnan, J., & Frauenfelder, U., (1988). Coreference assignment during sentence processing. In M. Macken (Ed.), *Language structure and processing.* Stanford, CA: CSLI.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41,* 67–85.

Wickelgren, W. A., Corbett, A. T. & Dosher, B. A. (1980). Priming and retrieval from short-term memory: A speed-accuracy tradeoff analysis. *Journal of Verbal Learning and Verbal Behavior, 19,* 387–404.