Language and Linguistics Compass 5/11 (2011): 764–783, 10.1111/j.1749-818x.2011.00313.x

Comprehension of Linguistic Dependencies: Speed-Accuracy Tradeoff Evidence for Direct-Access Retrieval From Memory

Stephani Foraker[1]* and Brian McElree[2]
[1]*Department of Psychology, SUNY College at Buffalo and* [2]*Department of Psychology, New York University*

Abstract

Comprehenders can rapidly and efficiently interpret expressions with various types of non-adjacent dependencies. In the sentence *The boy that the teacher warned fell*, *boy* is readily interpreted as the subject of the verb *fall* despite the fact that a relative clause, *that the teacher warned*, intervenes between the two dependent elements. We review research investigating three memory operations proposed for resolving this and other types of non-adjacent dependencies: *serial search retrieval*, in which the dependent constituent is recovered by a search process through representations in memory, *direct-access retrieval* in which the dependent constituent is recovered directly by retrieval cue operations without search, and *active maintenance* of the dependent constituent in focal attention. Studies using speed-accuracy tradeoff methodology to examine the full timecourse of interpreting a wide range of non-adjacent dependencies indicate that comprehenders retrieve dependent constituents with a direct-access operation, consistent with the claim that representations formed during comprehension are accessed with a cue-driven, content-addressable retrieval process. The observed timecourse profiles are inconsistent with a broad class of models based on several search operations for retrieval. The profiles are also inconsistent with active maintenance of a constituent while concurrently processing subsequent material, and suggest that, with few exceptions, direct-access retrieval is required to process non-adjacent dependencies.

Memory has long been recognized as an important determinant of language performance. Since Miller and Chomsky (1963), it is widely held that constraints on memory storage and retrieval of word representations and constituents limit the complexity of the expressions that comprehenders can successfully process. More recently, language scientists have come to appreciate that memory constraints not only determine the upper bound on complexity but also crucially affect the ease and accuracy of comprehending seemingly simple and common expressions. Whether interpreting spoken, written, or signed language, comprehenders must reconstruct linguistic relationships among the sequentially presented elements that encode meaning. To understand and model comprehension, then, a systematic investigation of the memory structures and operations used in real-time comprehension is needed.

This article reviews studies investigating the comprehension of different types of non-adjacent dependencies as a means of exploring the nature of the memory system supporting comprehension. For example, in the sentence *The boy with the green shirt that the teacher warned fell*, a memory representation of *boy* must be connected with the verb *fell* to interpret the sentence. In this review, the term ‘‘dependency’’ is used broadly, subsuming verb-argument(s), filler-gap, verb-phrase (VP) ellipsis, sluicing, as well as pronoun-referent dependencies. Exploring a range of dependencies tests the generality of the findings, and enables researchers to address whether the same type of operations are found at theoretically distinct levels of language processing.

Much of the theorizing about the role of memory in comprehension has been framed at levels that abstract away from the specific memory representations and operations used in comprehension. When memory constraints have been considered, they are typically simple memory structures (stacks, buffers, etc.) and principles (storage cost, decay, etc.) that are descriptive of the phenomenon rather than explanatory, and often lack strong independent psychological support (Lewis et al. 2006). The studies reviewed here take a different approach to the role of memory in comprehension: They are informed and motivated by principles derived from basic memory research, which specify explanatory mechanisms that have independent support, from outside the language system.

Following seminal work by Sternberg (1969), which reintroduced and extended Donders' subtraction method (1868/1969), timing measures of mental processing have been recognized as a preeminent way to address a host of questions concerning the fundamental properties and organization of mental architectures (Wickelgren 1977; Wickelgren and Corbett 1977). The key assumption underlying the application of such measures is that differences in processing time scale to differences in real-time mental operations. As a basic example of subtraction logic, if reading time for one sentence is longer than another, the extra time may reflect an additional processing step. When this assumption is valid, it licenses the use of timing measures to investigate ''architectural'' issues that are essential precursors to developing fully articulated models of language processing. Issues that timing measures can be employed to investigate include how component operations within a complex skill are organized (e.g., in serial, parallel, cascading), when certain parsing or interpretive processes are operative, and whether one source of information circumvents the use of another, among others. In this review, we do not lay out a specific model of sentence processing, but rather focus on the components and organization of the memory system supporting language comprehension.

The Memory System Underlying Comprehension

A traditional information-processing approach to the human memory system (e.g., Atkinson and Shiffrin 1968; Raaijmakers and Shiffrin 1981) consists of several components. The ones we will be most concerned with are *long-term memory*, an essentially unbounded, permanent repository for stored information that is passive (i.e., of which we are not currently aware), and short-term or *working memory*, a temporary, labile state for consciously manipulating or working with information.

Access to passively stored information in long-term memory is generally regarded as direct, via a cue-driven retrieval operation that activates stored representations (Clark and Gronlund 1996); an internet link that takes you to a particular page of content is an example of direct access. In language processing, stored information includes representations and knowledge about phonology, syntactic structures, semantic and pragmatic information, discourse and information structure, as well as knowledge about non-linguistic constraints such as frequency, transitional probabilities, etc. With an adequate set of retrieval cues, passive information is reactivated when needed for ongoing language processing.

On the other hand, access to information in working memory has traditionally been argued to involve serial processing, with a series of comparisons to all currently active concepts held in a limited-capacity storage component (Sternberg 1966, 1975; Theios 1973; Treisman and Doctor 1987); consider a set of tabs open on your browser, through which you must scan to find the one you want. Based on the ‘‘container’’ conception of working memory, comprehension errors or failures have been routinely attributed to overtaxing a limited working memory capacity when storing or manipulating the products of recent analyses (e.g., Caplan and Waters 1999; Fedorenko et al. 2006; Gibson 1998, 2000; Just and Carpenter 1992; King and Just 1991; MacDonald et al. 1992). However, the construct ‘‘capacity’’ lacks specificity, and the application of capacity-based approaches to different comprehension deficits has been criticized on several grounds (see MacDonald and Christiansen 2002). Notably, a growing body of research suggests that many of the comprehension errors or failures attributed to exceeding ‘‘resource limits’’ are more properly viewed as failures in retrieval stemming from retrieval interference (Gordon et al. 2001, 2002, 2004; Van Dyke 2002, 2007; Van Dyke and Lewis 2003; Van Dyke and McElree 2006).

With this issue in mind, the distinction between long-term vs. working memory does not map neatly to the memory operations discussed in this review. That is, whether ‘‘memory capacity’’ constrains comprehension in some fashion remains an open question. However, the question cannot be fully addressed without a detailed understanding of the memory operations in on-line comprehension, including how we retrieve representations, and what factors determine retrieval success.

CANDIDATE MEMORY OPERATIONS

Recent investigations of memory retrieval have identified two distinct ways in which information is retrieved, in *both* short-term and long-term domains (see McElree 2006). The first is *serial search retrieval*, which is a one-by-one, relatively slow search that appears to be necessary for recovering order information, such as the order of elements in time or across space (Gronlund et al. 1997; McElree 2001, 2006; McElree and Dosher 1993). Search may take several forms, which we introduce as appropriate later in the review. The second is *direct-access retrieval*, a one-step cue-driven operation which provides access to information about the item itself, via a content-addressable representation (McElree 1996, 1998, 2006; McElree and Dosher 1989, 1993; Öztekin and McElree 2007). Content-addressability means that cues at the retrieval site make contact with memory representations that have overlapping content, without recourse to a sequence of searches through irrelevant memories for the to-be-retrieved item (e.g., Clark and Gronlund 1996; Dosher and McElree 2003; Kohonen 1984). For example, the cues available at the point of retrieval ‘‘resonate’’ with items in memory according to the amount of partially matching content, and the item retrieved is the one with the most overlap or best fit (e.g., Ratcliff 1978).

We also consider *active maintenance* in focal attention to be a third fundamental cognitive operation involved in comprehension. Modern conceptions of the memory system include controlled attention, where one’s focus of attention is a very limited-capacity state into and out of which one shunts information very quickly. Several lines of evidence derived from a variety of cognitive and perceptual tasks indicate that only an extremely limited amount of information can be maintained in focal attention (3–4 units: Cowan 2001, 2005; 1 unit: McElree 1998, 2001, 2006). In this paper, we assume McElree’s (2006) conception of focal attention, stating that focal attention is just one processing chunk which is quickly replaced by the next chunk of information in mental processing. What defines the chunk of information does not appear to be a single ‘‘item,’’ but rather one processing epoch that forms one unitary chunk. For example, McElree (1998) sequentially presented 9-item lists consisting of three instances of three categories to participants for study (e.g., *dog*, *horse*, *pig*, *leg*, *head*, *foot*, *hat*, *dress*, *tie*). He found that all items from the most recent category (clothing) showed the same faster speed of retrieval compared to the other items and categories, indicating the last category of items formed one chunk that was actively maintained in focal attention.

DISTINGUISHING BETWEEN CANDIDATE OPERATIONS

The key property of a serial search is that retrieval time is affected by the number of items in the memory set that must be searched through prior to a response. As the number of irrelevant distractors in the set increases, retrieval time will slow (on average across trials). In contrast, for direct access, retrieval time is unaffected by increasing the number of distractors, since the retrieval process occurs only for the desired item that matches the cues. Hence, these two classes of models, serial search vs. direct access, can be contrasted with measures of retrieval speed. For serial search, retrieval speed will progressively slow as more distractors are searched through; for direct access, retrieval speed will be the same, no matter the number of distractors.

To determine whether information is actively maintained in focal attention, studies using several memory tasks provide clear evidence that processing speed is also an effective indicator. The logic is straightforward: information in focal attention should be processed at a quick, uniform speed, because it does not need to be retrieved before being brought to bear on on-going operations. Beginning with Wickelgren et al. (1980), studies of recognition memory have shown that if there is no other item or activity that intervenes between the study and test of an item, that item is recognized extremely quickly – from 30 to 50% faster than items at earlier, less recent positions on a study list (McElree 1996, 1998; McElree and Dosher 1989, 1993; for a review, see McElree 2006). In sum, processing speed is the lynchpin for evaluating which of the three candidate memory operations is involved during various comprehension circumstances. Serial search retrieval will slow with more distractors while direct-access retrieval will not, and direct-access retrieval will manifest as a uniformly slower speed than active maintenance in focal attention.

Unfortunately, linguistic expressions may differ in processing time for two independent reasons. One is that the *speed* of processing is actually slower: retrieving a constituent is slower, or interpreting the expression is slower, or both. The other is that the *accuracy* of processing is degraded: retrieving the desired constituent may incur errors (e.g., failure to retrieve essential information, misanalyses of grammatical relations), and/or the quality of the resulting interpretation is lesser (e.g., acceptability, plausibility, specificity). Crucially, common timing measures will be slower or later if there is decreased accuracy – even if all component operations required for interpreting each expression complete at comparable speeds (McElree and Nordlie 1999). This fact applies to common timing measures employed for studying language comprehension, such as reaction times (RTs) for judgments about an expression (e.g., acceptability, recognition, lexical decision), self-paced reading time, eye-movement measures while reading an expression, the timing and duration of gaze to objects in the ''visual world'' paradigm. As a consequence, researchers *cannot* straightforwardly infer an underlying difference in processing speed from a difference in commonly used timing measures, since the confounds of retrieval success and interpretation quality can affect processing time in addition to changes in retrieval/interpretation speed. What is required is a method that provides conjoint measures of speed and accuracy, enabling researchers to tease apart the contributions of the *speed* of retrieval and interpretation vs. the *accuracy* of retrieval and interpretation.

The Speed-Accuracy Tradeoff (SAT) Procedure

One experimental procedure that provides conjoint measures of the quality of information processing and the speed with which that processing takes place is the response-signal SAT procedure. This method of modeling the contributions of speed and accuracy has figured prominently in research investigating the memory system's architecture (Dosher, 1979; Reed 1973, 1976; Wickelgren 1977; Wickelgren and Corbett 1977), so we begin with examples from basic memory research.

"Speed-accuracy tradeoff" refers to the observation that one can perform a task at a faster speed but with the cost of lower accuracy, or alternatively at a slower speed but with higher accuracy. In the SAT procedure, this tradeoff is partially controlled by having participants respond at set time points. As quickly as possible after a cross-modal signal, the participant makes a response. The signals can be distributed across trials with one deadline per trial (single-response SAT), or they can all occur on each trial (multiple-response SAT). Signal time points are chosen to capture the full span of processing during a task, from before the beginning until after the end. In this way, the procedure provides a window onto mental processing as it unfolds over time, systematically probing the development of the processing in question.

Response accuracy for each time point is calculated as d' (to control response bias), which is the hit rate minus the false alarm rate. As a memory task example, a hit is an item correctly remembered (you remember *bed*, which was studied), while a false alarm is an item incorrectly remembered (you remember *sleep*, but it was not studied).

Hypothetical data shown in Figure 1 illustrate accuracy (black circles) after different amounts of processing time for one condition. A curve representing the growth of response accuracy is also shown (solid line), fit with an exponential approach to a limit, plotting accuracy (d') as a function of processing time (t): $d' = \lambda\ (1 - e^{-\beta(t-\delta)})$ for $t > \delta$, otherwise 0.

Speed-accuracy tradeoff functions have three phases. They first show an initial period of chance performance ($d' = 0$), followed by a monotonically increasing function, culminating in a final asymptotic level. The *intercept* of the function, estimated by the parameter δ, represents the point in time when accuracy departs from chance. The *rate* at which accuracy grows from chance to asymptote, estimated by β, measures the rate of information accrual. Jointly, intercept and rate measure of the *speed* of processing, indexing how quickly accuracy accrues to its asymptotic level. The *asymptote* of the function, λ, is a measure of terminal accuracy, providing an estimate of the highest level of *accuracy* reached with maximal processing time.

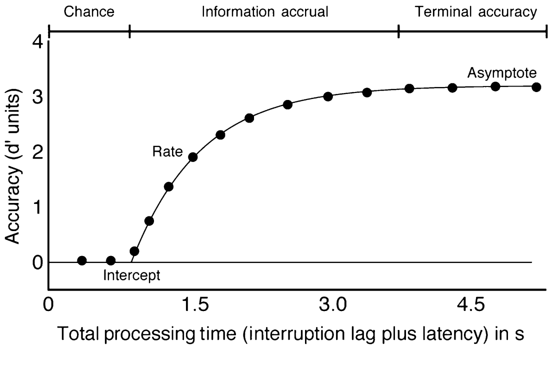

Fig 1. An speed-accuracy tradeoff (SAT) function for one condition, illustrating the three phases of processing.

To find the best fitting set of parameters for different conditions, the number of asymptotes, intercepts, and rates are systematically varied from a null model (one asymptote, rate, and intercept, 1λ-1β-1δ, for all data points from all conditions), through hypothesis-driven combinations, to a fully saturated model, such as 4λ-4β-4δ for four conditions. The best fitting model (see Figures 2–5) is chosen by a combination of criteria. All of the SAT publications noted below provide more technical details.

Examples From Basic Memory Research

SERIAL SEARCH VS. DIRECT-ACCESS RETRIEVAL

Figure 2 presents data from an experiment of McElree and Dosher (1993) showing the pattern characteristic of serial search: the rates and/or intercepts slow as a function of the increasing number of items that must be searched through in memory. The SAT functions are from a judgment of recency task, in which a list of six items (Q B F L X K) is sequentially presented, followed by two items from the list (Q K) presented together. Participants performed two-alternative forced-choice recency discriminations (was Q or K seen more recently?) which required recovering relational order information to choose the more recent item (K). Plotted are test pairs (e.g., sp16) composed of the first serial position item on the list (sp1 = Q) paired with items from other serial positions (sp6 = K, etc.).

As an item appeared less recently, asymptotic accuracy decreased (6 > 5 > 4 > 3 > 2). Crucially, there was *also* systematic slowing of retrieval speed: intercepts were progressively later in time and rates slowed (flatter slope) as the item decreased in recency. Model fits showed that this difference is inconsistent with direct-access retrieval (or parallel search models), and that it implicates a backwards serial search, where a memory representation of the list is searched from the last item studied back through the list.

In contrast, Figure 3 shows the pattern characteristic of direct-access retrieval: the rates and/or intercepts are not affected by increasing the number of items studied. In McElree (1996), participants studied five items presented sequentially, and then performed a yes–no item recognition task (was this item on the list?). The recognition probe consisted of either an item on the list or an unstudied item, which required information about the item's identity, but not order or relational information (i.e., recency).

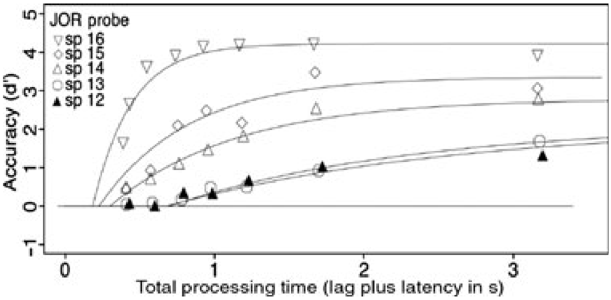

Fig 2. Speed-accuracy tradeoff (SAT) functions from a Judgment of Recency task showing the characteristic slowing of retrieval speed (intercept and/or rate) predicted for a serial search; best fit model 5λ-5β-4δ. (Plot from data in McElree and Dosher 1993.)

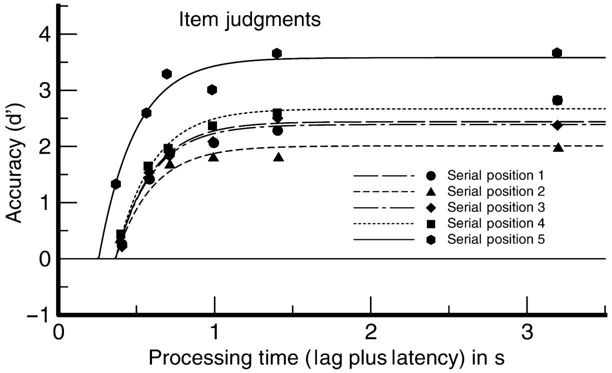

Fig 3. Speed-accuracy tradeoff (SAT) functions from a yes–no item recognition task showing the characteristic pattern of identical retrieval speed (intercept and rate), indicating direct-access retrieval for all items outside the focus of attention (Serial Positions 2–5); best fit model 5λ-1β-2δ. (Plot reproduced with permission from McElree 1996.)

As with judgments of recency (Figure 2), asymptotic accuracy systematically decreased as the test probe was from less recent positions (5 > 4 > 3/1 > 2). Critically, though, recency did not systematically affect retrieval speed for list items 1–4, which were accessed at the same speed (same rate and intercept). (See next section for discussion of the exception, the most recent item in Serial Position 5.) This pattern is representative of all SAT investigations of the effects of recency and memory list length on the retrieval of item information (e.g., McElree and Dosher 1989, 1993; McElree 1996, 1998, 2006; Öztekin and McElree 2007; Wickelgren et al. 1980), which have consistently demonstrated that neither the recency of the test probe nor the size of the memory list affects retrieval speed. The lack of a list length effect on retrieval speed is inconsistent with an exhaustive search of memory, whether serial (Sternberg 1966, 1975; Treisman and Doctor 1987) or parallel (Ratcliff 1978), and the lack of a recency effect is inconsistent with self-terminating searches (Murdock 1971; Theios 1973; Townsend and Ashby 1983). Collectively, the consistent speed exhibited by SAT modeling indicates that participants have direct access to item representations, even when those representations differ in quality (viz., engender lower asymptotes).

DIRECT-ACCESS RETRIEVAL VS. ACTIVE MAINTENANCE

Figure 3, above, also shows results demonstrating the signature speed advantage that is diagnostic of actively maintaining an item in focal attention. A test probe that is the last item on the study list, Serial Position 5, is the case where no other item intervenes between study and test. This condition is associated with exceptionally fast processing dynamics (rate or intercept) compared to test probes from all other positions. Since no other processing occurs that would displace the last item from focal attention, it is reasonable to attribute the observed advantage to uninterrupted matching of the test probe to the contents of focal attention.

Several lines of evidence indicate that the speed advantage for the most recent item observed in McElree (1996, 2001) and other SAT studies is uniquely linked to focal attention. In addition to fast matching of the most recent item with its adjacent test probe, procedures that encourage subjects to reinstate items from different list positions into focal attention provide the most direct support for this claim. McElree (2006) found a speed advantage for items from a less recent part of the list when participants were precued to retrieve those items just before a test, so they were already active at test. Further, in a controlled rehearsal study, in which participants were instructed to covertly rehearse items on the list when given an external signal, McElree (2006) found that the advantage tracked with the items participants were assumed to be actively rehearsing at test time.

Importantly, clear correlates of this speed advantage are found in other timing measures. In RT tasks, items associated with fast SAT dynamics show the expected shorter mean RTs, but more importantly, the entire RT distributions are shifted toward earlier times (McElree, 1993; McElree 1998; Oberauer 2002, 2006; Verhaegen et al. 2004). In addition, in recent fMRI studies (Öztekin et al. 2008, 2010), conditions that engendered a fast SAT speed were associated with less activity in the hippocampus, a region involved with successful episodic retrieval, and less activity in the inferior frontal gyrus, a region thought to be involved with retrieval effort (e.g., Cabeza et al. 2002). This pattern suggests that retrieval operations are not required for information currently active in focal attention.

Memory Operations Underlying Comprehension

SERIAL SEARCH VS. DIRECT ACCESS

Which types of memory operation are used in language comprehension and under what circumstances? Inasmuch as the hierarchical structure of a sentence is often encoded by the order of constituents within a string, predominantly so in languages such as English, one could argue that a serial search like that used to retrieve relational (e.g., recency) information might be best for accessing the elements involved in non-adjacent dependencies. However, many parsing models assume that representations are content-addressable and accessed without a search (Stevenson 1994; Tabor and Hutchins 2004; Vosse and Kempen 2000). SAT studies of resolving non-adjacent dependencies provide the required empirical support for this assumption.

Relevant experiments have employed the same logic used to investigate basic memory retrieval operations. To contrast serial search and direct-access retrieval, increasing amounts of material are placed in the supposed ''search path'' between the two dependent elements. The key prediction is that the speed of resolving a dependency will systematically slow with longer search paths, with differences analogous to those in Figure 2. Alternatively, if additional interpolated material does not affect the speed of resolving the non-adjacent dependency, direct-access retrieval is supported, analogous to Figure 3 (Serial Positions 1–4).

To explain how the SAT procedure has been adapted for investigating memory operations involved in language comprehension, we first describe one study in more detail. McElree (2000) reported the first study investigating whether material interpolated between a verb and its argument affected comprehension speed. In examples (1–3), the direct object noun phrase (NP, *the book*) of the matrix verb (*admired*) was fronted to the beginning of the sentence in a cleft construction, creating a filler-gap dependency. The retrieval site at which the dependency needs to be resolved is shown in bold, in these cases, the final verb.

(1) This was the book that the editor **admired**.

(2) This was the book that the editor who the receptionist married **admired**.

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

The distance between the NP and verb was increased by adding one (2) or two (3) subject-relative clauses. Participants judged the acceptability of the sentences, which required discriminating acceptable ones like (1–3) from unacceptable counterparts, where the verb *amused* replaced *admired*. A hit is when ''the editor admired the book'' is correctly judged acceptable, and a false alarm is when ''the editor amused the book'' is incorrectly judged acceptable. The source of unacceptability in comprehension experiments is carefully chosen to tap the dependency in question, at an appropriate level of processing (e.g., violating grammatical selectional restrictions of the verb, not a spelling/phonology error). In addition to acceptable and unacceptable experimental materials, control sentences that guard against shallow or incomplete processing strategies are included.

This experiment used the single-response SAT procedure: on each trial, participants read a sentence word-by-word (i.e., rapid serial visual presentation) and entered an un/acceptable button press judgment for one of six response-signals ranging from 50 to 3000 ms after the onset of the final verb. They responded as quickly as possible within a 300 ms window, for which they had previous training.[1]

Analyses of the SAT functions indicated that the asymptotes decreased progressively with more interpolated material. This pattern indicates a progressively lower likelihood of retrieving the correct argument from memory and computing a correct interpretation of the sentences. When interpolated material causes interference during direct-access retrieval, lower accuracy is predicted (for the cue *admired* vs. **amused*, **receptionist* interferes with *book*). However, the speed of comprehension (rate/intercept) was *not* affected by the amount of material intervening between the dependent elements, a pattern consistent with direct access and at odds with a search operation.

McElree et al. (2003) extended the investigation further by exploring the effect of a different type of interpolated structure and different types of dependencies. In one experiment, embedded complement clauses were inserted between a clefted NP (*the scandal*) and the final verb (*panicked/*relished*), as shown in (4–6).

(4) It was the scandal that the celebrity **relished**.

(5) It was the scandal that the model believed that the celebrity **relished**.

(6) It was the scandal that the model believed that the journalist reported that the celebrity **relished**.

Embedded complement clauses differ from center-embedded subject-relative clauses in that they increase not only the surface distance between the verb and its argument but also the distance along the right edge of a hierarchical structure (see McElree et al. 2003). Contrasts such as (4–6) provide a test of whether hierarchical distance rather than surface distance is a determinant of search time. Figure 4 shows the average SAT functions for the contrasts in examples (4–6). Like McElree (2000), accuracy declined as the distance between the dependent elements increased, but the speed of processing remained constant.

They also examined subject-verb dependencies, contrasting adjacent subject (*book*) and verb (*ripped/*laughed*; 7) with cases where different structures intervened: (8) object relative clause, (9) prepositional phrase and object relative clause, (10) object relative and subject-relative clause, or (11) two object-relative clauses.

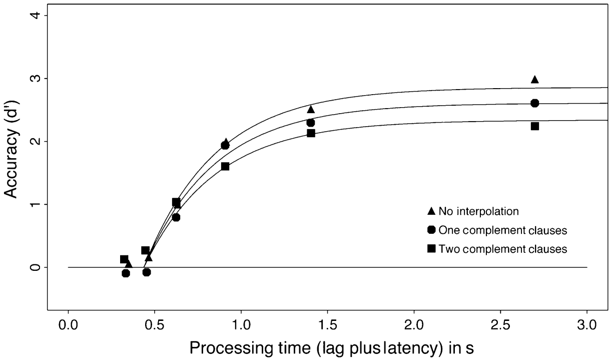

Fig 4. Average speed-accuracy tradeoff (SAT) functions for the contrasts in examples (4–6); best fit model 3λ-1β-1δ. (Plot reproduced with permission from McElree et al. 2003.)

(7) The book **ripped**.

(8) The book that the editor admired **ripped**.

(9) The book from the prestigious press that the editor admired **ripped**.

(10) The book that the editor who quit the journal admired **ripped**.

(11) The book that the editor who the receptionist married admired **ripped**.

For (7–11), asymptotic accuracy decreased systematically with the amount of interpolated material, again suggesting retrieval interference. Crucially, the speed of processing was the same for (8–11). Speed did not systematically slow as the amount of interpolated material increased, consistent with McElree (2000) and Figure 4. Only when the verb was adjacent to its subject (7) was speed faster, indicating that the last item processed was still active in focal attention. We discuss this result further, below.

A final experiment examined "tough" constructions, again varying the distance of the search path. Shown below is the dependency between a verb particle (*spread open*) and its direct object NP (*the album*), with short (12) and long (13) distances between.

(12) This is the album that the customer found difficult to **spread open**.

(13) This is the album that the customer who obviously angered the fussy collector found difficult to **spread open**.

In addition, double-argument dependencies were included to test whether search was used when multiple bindings at the verb were critical for comprehension. Memory research indicates that a serial search is required only when relational or order information is involved in the dependency (McElree 2006). Thus, direct-access retrieval may be insufficient when constructing an interpretation that depends on the relations of multiple constituents (direct vs. indirect object). In (14) and (15), the verb particle (*mount in*) required correctly binding two arguments (*album* and *stamps*) to construct the ditransitive verb phrase (*mount the stamps in the album/#mount the album in the stamps*).

(14) This is the album that the stamps were difficult to **mount in**.

(15) This is the album that the stamps which obviously angered the fussy collector were difficult to **mount in**.

Figure 5 shows the SAT functions. Note that for both single- and double-argument sentences, additional material (short vs. long) only lowered the asymptotes. Distance did not affect the timecourse parameters, and did not interact with the number of dependencies (single vs. double). The difference in timecourse seen is due only to the number of arguments, with double-argument structures resolved at a slower rate than single-argument structures. When multiple dependencies need to be resolved at a gap, which may require using relational information, a slower search operation seems to be used.

Broadening the type of dependency considered, recent work has examined how an antecedent is reaccessed during VP ellipsis (Martin and McElree 2008, 2009) and sluicing (Martin and McElree 2011). Like other non-adjacent dependencies, ellipsis and sluicing require access to the representation of a previously processed constituent to interpret the elided constituent or *wh*-remnant. However, they differ from filler-gap and subject-verb dependencies in an important respect. Comprehenders could predict that a dislocated NP or the subject of a matrix clause will be needed at a later point in the sentence. Indeed, parsing models often assume that these types of constituents are held in specialized stacks or buffers, and some of these mechanisms can mimic properties of a direct-access operation (McElree et al. 2003). This is not true of an antecedent for ellipsis or sluicing. In both cases, the antecedent is already fully integrated with its local context, reducing strategic expectancies that it will be needed later on. As such, VP-ellipsis and sluicing are critical test cases for investigating whether all representations formed in comprehension are retrieved with a direct-access operation.

Martin and McElree (2008, 2009) tested for differences in the speed and accuracy of processing the ellipsis site, *did not*, as a function of its distance from the antecedent, *admired the author's writing*.

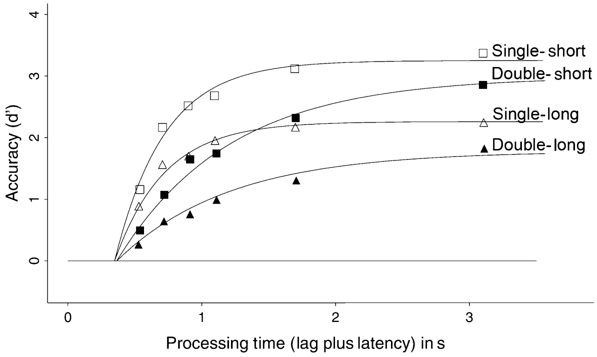

Fig 5. Average speed-accuracy tradeoff (SAT) functions for single-argument sentences with short (12) or long (13) distance, and double-argument sentences with short (14) or long (15) distance; best fit model 4λ-2β-1δ. (Plot reproduced with permission from McElree et al. 2003.)

(16) The editor *admired the author's writing*, but the critics **did not**.

(17) The editor *admired the author's writing*, but everyone at the publishing house was shocked to hear that the critics **did not**.

Martin and McElree (2008) varied the distance between antecedent and ellipsis in various ways, one of which is shown in (17). Martin and McElree (2009) further tested for different types of serial search. For *forward search* from the beginning of the sentence, the amount of material before the antecedent was varied, and for *backward search* from the ellipsis site, the amount of material between antecedent and ellipsis was varied. In some cases, the type of intervening material decreased asymptotic accuracy, consistent with interference between content-addressable retrieval cues. However, the speed of interpreting the ellipsis was not affected in any of the six experiments, including the experiment that tested for forward and backward searches, supporting the direct-access retrieval account.

Their 2011 investigation examined sluicing, which is an important dependency to consider because it allows a comparison of *syntactically guided* search vs. direct-access retrieval. The alternative of syntactically guided search proposes that search occurs only through syntactically licensed entities. This is a potential issue for the evidence presented so far, because syntactic information could effectively limit the set of possible fillers/antecedents to just the one that fulfills the non-adjacent dependency; search through a ''list'' of one would produce the same results as direct-access retrieval. Martin and McElree therefore manipulated the number of syntactically available antecedents (18 and 19 one, 20 and 21 two), as well as the distance between antecedent, *studied*, and sluice site, *what* (18 and 20 recent, 19 and 21 distant).

(18) In the morning, Michael *studied* but he didn't tell me **what**.

(19) Michael *studied* in the morning, but he didn't tell me **what**.

(20) Michael slept and *studied*, but he didn't tell me **what**.

(21) Michael *studied* and slept, but he didn't tell me **what**.

In (20) and (21), both the correct antecedent *studied* and the incorrect verb *slept* are syntactically licensed. If search is syntactically constrained, the presence of *slept* should slow the speed of interpretation compared to (18) and (19). As well, if syntactically guided search occurs in a forward fashion, then (20) should be slower than (21), and vice versa if it is backward. The results, however, revealed no difference in the speed of processing, contra syntactically guided search, either forward or backward. Instead, the asymptotic differences supported direct-access retrieval.

DIRECT-ACCESS RETRIEVAL VS. ACTIVE MAINTENANCE

Much less research has contrasted direct-access retrieval with active maintenance during comprehension. To date, the focus has been on the preliminary yet theoretically important issue of the effective span of focal attention: what constitutes a chunk that can be actively maintained over intervening material? Determining the effective span of focal attention in sentence comprehension is an important and challenging question, because linguistic relations provide a basis for forming richly structured representations that could form a unitary chunk. Yet, most results in SAT comprehension studies have shown a speed advantage analogous to results from memory research: Comprehension is faster when two dependent constituents are adjacent than when other constituents intervene (Foraker 2007; Foraker and McElree 2007; McElree et al. 2003; Wagers and McElree forthcoming). As discussed above, McElree et al. (2003) reported a faster speed of comprehension for an adjacent subject-verb dependency (7) than for conditions in which different types of relative clauses intervened between the subject and verb (8–11).

Wagers and McElree (forthcoming; Wagers 2010) have explored systematically whether all types of sentential elements serve to displace a subject phrase from focal attention. With constructions such as those in (22–26), they used SAT modeling to investigate under what conditions the subject NP *the driver* was no longer active in focal attention at the matrix verb *fainted* (**drained*).

(22) The crowd gasped as the driver **fainted**.

(23) The crowd gasped as the driver of the ambulance **fainted**.

(24) The crowd gasped as the driver who wrecked the ambulance **fainted**.

(25) The crowd gasped as the driver who the ambulance hit **fainted**.

(26) The crowd gasped as the driver abruptly **fainted**.

Results indicated that subject NPs can be maintained in focal attention concurrently with either a modifying prepositional phrase (23) or an adverb (26), exhibiting the same time-course (intercept, rate) as the bare NP in (22). The relative clauses in (24) and (25), however, displaced the subject, showing a slower speed of comprehension. These results demonstrate that the capacity of focal attention is sensitive to linguistic structure, but nevertheless extremely limited.

A related issue also in preliminary stages of investigation is whether linguistic structure and devices for focusing information modulate active maintenance of a constituent. From SAT memory experiments, we know that participants can strategically maintain (McElree 2001) or pre-activate (McElree 2006) an item while continuing to process intervening material, demonstrating that predictive, active maintenance is possible. Applied to language comprehension, this approach is in the spirit of a strong interpretation of classic psycholinguistic studies supporting the *Active Filler Strategy* when resolving dependencies (Crain and Fodor 1985; Fodor 1978; Frazier 1987; Frazier and Clifton 1989; Frazier and Flores d'Arcais 1989; Stowe 1986; see also Wanner & Maratsos' (1978) ''hold'' mechanism). The principle is that as soon as a filler has been identified, the parser will actively predict potential gap sites as the comprehender progresses through a sentence (Frazier and Clifton 1989). We interpret this strategy to involve active maintenance of the filler element as long as the dependency remains unresolved. The strategy's critical aspect that we tested is whether proactive prediction rather than reactive retrieval is a psychologically plausible operation during comprehension.

Foraker and McElree (2007) investigated whether clefting structures engage such a predictive strategy during pronoun resolution (see also Gundel 1999). Clefting makes a referent prominent in a discourse, indicated by decreased reading time at the co-referring pronoun (Foraker 2004) or NP (Almor 1999). But is the reading advantage due to active maintenance (faster speed) or direct-access retrieval (higher accuracy)? In a series of SAT experiments, we manipulated the referent's prominence as well as its distance from the pronoun. In (27) and (30), the referent *boyfriend* is more prominent (underlined), while in (28) and (29) *ring* is more prominent. The referent and pronoun are adjacent in (27) and (28), and non-adjacent in (29) and (30).

(27) The one whom the engagement ring impressed was the ardent *boyfriend*. **He** stared/*sparkled.

(28) It was the engagement ring that impressed was the ardent *boyfriend*. **He** stared/*sparkled.

(29) It was the engagement *ring* that impressed the ardent boyfriend. **It** sparkled/*stared.

(30) The one whom the engagement *ring* impressed was the ardent boyfriend. **It** sparkled/*stared.

Overall, we found that when the pronoun referred back to a prominent referent, asymptotic accuracy was higher than the non-prominent conditions, consistent with facilitated retrieval of the referent representation. However, prominence did not affect the speed of processing. Instead, distance affected the speed of retrieval and resolution, with a faster rate for adjacent compared to non-adjacent elements.

Similar results have been found for prominence as a result of spoken prosodic stress during listening (Foraker 2007). Prosodic stress increased the accuracy of retrieval/resolution but did not affect the speed. On the other hand, comprehension occurred at a faster speed when the referent and pronoun were adjacent, compared to when they were non-adjacent.

Discussion

We have reviewed SAT studies investigating the timecourse of processing several types of non-adjacent dependencies: subject-verb and filler-gap dependencies, VP ellipsis, sluicing, and pronoun co-reference. Although different constraints govern each type of dependency, the evidence indicates that access to a previously processed dependent constituent is direct. In particular, the unvarying speed of comprehension (consistent rates and/or intercepts) across different distances between dependent elements supports this type of operation. The timecourse evidence is inconsistent with search operations, including forward (Martin and McElree 2009, 2011; see also Van Dyke and McElree 2007) and backward serial search (Martin and McElree 2008; McElree 2000; McElree et al. 2003; Wagers and McElree, forthcoming), and syntactically guided search (Martin and McElree 2011). To date, only when multiple dependencies are resolved at a gap, which relies on relational information, does a search operation seem to be used (McElree et al. 2003).

Evidence in hand also indicates that the representations formed during comprehension are content-addressable, which allows them to be accessed directly by retrieval cues at the dependency site. A notable advantage of a direct-access memory system is that it enables the rapid recovery of past representations, without introducing the distance-dependent processing time cost found for search. To a large degree, the rapidity of language processing may be a consequence of the use of cue-driven, direct-access operations. On the other hand, the disadvantage of a direct-access memory system is that it is highly susceptible to interference at retrieval, which arises when other constituents in memory have properties matching the cues used for retrieval. Basic memory research indicates that retrieval interference results from *cue-overload*: retrieval cues cannot reliably elicit a desired target because they are associated with other items in memory (e.g., Nairne 2002; Öztekin and McElree 2007; Watkins and Watkins 1975), and this reduces the quality or distinctiveness of the desired representation. In language comprehension, failing to retrieve a required constituent or retrieving the wrong constituent would result in a degraded or even anomalous interpretation, correctable only by reanalysis. The nearly ubiquitous differences in asymptotic accuracy observed in the reviewed studies are readily explained by this type of retrieval interference: increasing amounts of additional material placed before or between a non-adjacent dependency progressively reduces the probability of retrieving the required constituent, thereby reducing the probability of deriving an acceptable interpretation of the longer expressions.

Current evidence also indicates that the span of focal attention in comprehension is extremely limited. The rich structure of linguistic relations does not, by in large, provide a basis for forming a representation that can be actively maintained (McElree et al. 2003; Wagers and McElree, forthcoming).

Consequently, retrieval operations are likely required to establish relations between most dependent constituents that are not adjacent to one another in the input stream. Even simple expressions like unambiguous relative clauses may require shunting information between memory and focal attention. Similarly, linguistic devices such as clefting and prosodic stress that make a referent prominent do not result in active maintenance of the referent representation (Foraker 2007; Foraker and McElree 2007). Although such devices do facilitate resolving a dependency, the effects are because of retrieval operations, rather than a predictive ''maintain'' operation.

Broader Implications

These findings indicate that memory operations in comprehension function according to principles at work in other domains, informing the longstanding debate about whether or not language has a dedicated working memory system (e.g., Caplan and Waters 1999; MacDonald and Christiansen 2002). The studies discussed here support content-addressable, direct-access retrieval, with retrieval cues providing all the specificity of function required for comprehension (e.g., morphosyntactic, semantic, and pragmatic constraints). Collectively, the evidence that comprehension relies on cue-driven, direct-access memory operations (Martin and McElree 2008, 2009, 2011; McElree 2000; McElree et al. 2003; Wagers and McElree, forthcoming) and the evidence that retrieval interference has adverse effects on language comprehension (e.g., Gordon et al. 2001, 2002, 2004, 2006; Van Dyke 2002, 2007; Van Dyke and Lewis 2003; Van Dyke and McElree 2006) presents an alternative to traditional accounts of comprehension errors or failures as arising from overtaxing a limited ''resource capacity.'' This alternative should be favored because it is both empirically grounded and explicit enough to be incorporated into computational models of sentence processing (e.g., Lewis & Vasishth 2005; Lewis et al. 2006; see also Lewis 1996). However, it is certainly possible that common operating *principles* might be used in otherwise distinct systems, rather than relying on one, common *mechanism* across systems. More compelling evidence would consist of demonstrations that the memory operations in comprehension also recruit the same brain regions as those outside the language domain (e.g., Öztekin et al. 2008, 2010), an issue for future research.

Marcus (2004) has suggested that a cue-driven, content-addressable memory system is suboptimal for language from a design perspective. Marcus asserts that location-based rather than content-based retrieval is the optimal design for language because it is not adversely affected by similarity-based interference or distance.

We do not share Marcus's intuitions that location-based retrieval is optimal for language. Even in languages with minimal case marking, such as English, location in a hierarchical structure only serves to uniquely identify some types of dependent constituents, and even strong structural constraints can sometimes be overridden (c.f., reflexives vs. logophors). Location-based structural constraints are weak to non-existent in many types of dependencies (e.g., VP ellipsis, sluicing, co-reference relations). For Marcus's argument to go forward, languages would need to disallow all dependencies other than those that are constrained by location, including those that are determined by case marking, which would substantially reduce expressive power (Fodor 1978). Crucially, location-based retrieval may be far less efficient than content-based retrieval in dealing with temporary structural ambiguities, which are ubiquitous in natural language. Unlike a content-based operation, location-based retrieval cannot exploit disambiguating information (e.g., lexical, semantic, pragmatic constraints) often available at the retrieval site.

Likewise, we do not share the intuition that a content-addressable memory system is suboptimal. Marcus does not specify the criteria he believes are relevant for assessing optimality, citing only the tendency for a content-addressable system to be susceptible to interference and distance effects. Indeed, some have argued that sensitivity to recency (distance) is in fact a rational property of a memory system (Anderson 1990). We agree that sensitivity to recency is not an intrinsic or necessary property of a content-addressable system (nor a location-based system, either). However, susceptibility to interference *is* an inherent property of a content-addressable system, and indirectly facilitates a host of subsequent operations that are sensitive to recency, such as ''priming'' effects (Dosher and Rosedale 1997; Ratcliff and McKoon 1994). Questions of optimality must weight costs against benefits over a circumscribed domain, taking into account relevant biological constraints. We do not believe our current understanding of the memory operations deployed during comprehension enables an informed assessment of the optimality of a content-addressable system. However, given that we have presented abundant evidence supporting content-addressable, direct-access retrieval, but also limited cases of a search-like process for resolving order relations and a fast active maintenance operation for adjacent dependencies, we expect that the optimal design for real-time comprehension will be a memory system with richly structured representations that can be accessed with different operations given different language structures.

Short Biographies

Stephani Foraker's research investigates the memory mechanisms involved in understanding language, focusing on co-reference dependencies. She has co-authored papers appearing in *Journal of Memory and Language* and *Cognitive Science*. She received her PhD (Experimental Psychology) at New York University, and is an Assistant Professor of Psychology at SUNY College at Buffalo.

Brian McElree has pioneered the use of speed-accuracy tradeoff modeling to investigate language processing issues. He has authored or co-authored numerous papers addressing the cognitive structures and processes that enable language comprehension, as well as more general issues concerning basic mechanisms in working memory and visual attention. McElree has served on the editorial board of *Journal of Memory and Language* as well as numerous grant panels. He received his PhD (Experimental Psychology) at Columbia University, and is Professor of Psychology at New York University.

Acknowledgement

We thank Lyn Frazier, another reviewer and the editors for helpful comments. Portions of the research reviewed were supported by NIH/NICHD grant R01-HD-056200 to B. McElree.

Notes

* Correspondence address: Stephani Foraker, Department of Psychology, SUNY College at Buffalo, 1300 Elmwood Ave., Buffalo, NY 14222, USA. E-mail: forakesm@buffalostate.edu

[1] In the multiple-response SAT procedure, several signals occur on each trial, typically spaced 300 ms apart, starting just before the un/acceptable point to 3000 ms later. Participants are trained to respond rhythmically to each signal, modulating their response to reflect their current judgment of the sentence.

Works Cited

Almor, Amit. 1999. Noun-phrase anaphora and focus: the informational load hypothesis. Psychological Review 106. 748–65.

Anderson, John R. 1990. The adaptive character of thought. Hillsdale, NJ: Erlbaum.

Atkinson, R. C., and Richard M. Shiffrin. 1968. Human memory: a proposed system and its control processes. The psychology of learning and motivation: advances in research and theory, vol. 2, ed. by K. W. Spence and J. T. Spence, 89–195. New York: Academic Press.

Cabeza, Roberto, R. Florin Dolcos, Reiko Graham, and Lars Nyberg. 2002. Similarities and differences in the neural correlates of episodic memory retrieval and working memory. Neuroimage 16. 317–30.

Caplan, David, and Gloria S. Waters. 1999. Verbal working memory and sentence comprehension. Behavioral and Brain Sciences 22. 77–126.

Clark, Steven E., and Scott D. Gronlund. 1996. Global matching models of recognition memory: how the models match the data. Psychonomic Bulletin & Review 3. 37–60.

Cowan, Nelson. 2001. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. Behavioral and Brain Sciences 24. 87–185.

——. 2005. Working memory capacity (Essays in cognitive psychology). New York: Psychology Press.

Crain, Stephen, and Janet D. Fodor. 1985. How can grammars help parsers. Natural language parsing: psycholinguistic, computational, and theoretical perspectives, ed. by D. Dowty, D. Kartunnen and A. M. Zwicky, 94–128. Cambridge, UK: Cambridge University Press.

Donders, F. C. 1969. On the speed of mental processes. Attention and performance II, ed. and trans. by W. G. Koster. Amsterdam: North-Holland (Original work published in 1868).

Dosher, Barbara A. 1979. Empirical approaches to information processing: speed–accuracy tradeoff or reaction time. Acta Psychologica 43. 347–59.

——, and Brian McElree. 2003. Memory Search. Learning and Memory, 2nd edn., ed. by John H. Byrne, 373–9. New York, NY: Macmillan Reference USA. Gale Virtual Reference Library.

——, and Glenda S. Rosedale. 1997. Configural processes in multi-cue priming of recognition. Cognitive Psychology 33. 209–67.

Fedorenko, Evelina, Edward Gibson, and Doug Rohde. 2006. The nature of working memory capacity in sentence comprehension: evidence against domain-specific working memory resources. Journal of Memory and Language 54. 541–53.

Fodor, Janet D. 1978. Parsing strategies and constraints on transformations. Linguistic Inquiry 9. 427–73.

Foraker, Stephani. 2004. The mechanisms involved in the prominence of referent representations (Doctoral Dissertation, New York University). UMI ProQuest Digital Dissertations.

——. 2007. Explicit versus implicit prosody: effects on pronoun interpretation. Poster presented at the CUNY Sentence Processing Conference, March 29–31, La Jolla, CA.

——, and Brian McElree. 2007. The role of prominence in pronoun resolution: active versus passive representations. Journal of Memory and Language 56. 357–83.

Frazier, Lyn. 1987. Syntactic processing: evidence from Dutch. Natural Language and Linguistics Theory 5. 519–60.

——, Charles Clifton, Jr. 1989. Successive cyclicity in the grammar and the parser. Language and Cognitive Processes 4. 93–126.

——, and Giovanni B. Flores d'Arcais. 1989. Filler-driven parsing: a study of gap filling in Dutch. Journal of Memory and Language 28. 331–44.

Gibson, Edward. 1998. Linguistic complexity: locality of syntactic dependencies. Cognition 68. 1–76.

——. 2000. The dependency locality theory: a distance-based theory of linguistic complexity. Image, language, brain: papers from the first mind articulation project symposium, ed. by A. Marantz, 94–126. Cambridge, MA: MIT Press.

Gordon, Peter C., Randall Hendrick, and Marcus Johnson. 2001. Memory interference during language processing. Journal of Experimental Psychology: Learning, Memory and Cognition 27. 1411–23.

——, ——, and ——. 2004. Effects of noun phrase type on sentence complexity. Journal of Memory and Language 51. 97–114.

——, ——, ——, and Yoonhyoung Lee. 2006. Similarity-based interference during language comprehension: evidence from eye tracking during reading. Journal of Experimental Psychology: Learning, Memory and Cognition 32. 1304–21.

——, ——, and William H. Levine. 2002. Memory-load interference in syntactic processing. Psychological Science 13. 425–30.

Gronlund, Scott D., Mark B. Edwards, and Daryl D. Ohrt. 1997. Comparison of the retrieval of item versus spatial position information. Journal of Experimental Psychology: Learning, Memory, and Cognition 23. 1261–74.

Gundel, Janet K. 1999.On different kinds of focus. Focus: linguistic, cognitive, and computational perspectives, ed. by P. Bosch and R. van der Sandt, 183–97. Cambridge: Cambridge University Press.

Just, Marcel A., and Patricia A. Carpenter. 1992. A capacity theory of comprehension: individual differences in working memory. Psychological Review 99. 122–49.

King, Jonathan, and Marcel Just. 1991. Individual differences in syntactic processing: the role of working memory. Journal of Memory and Language 30. 580–602.

Kohonen, Teuvo. 1984. Self-organization and associative memory. New York, NY: Springer.

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

——, and Shravan Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. Cognitive Sciences 29. 375–419.

——, ——, and Julie A. Van Dyke. 2006. Computational principles of working memory in sentence comprehension. Trends in Cognitive Sciences 10. 447–54.

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

——, Marcel A. Just, and Patricia A. Carpenter. 1992. Working memory constraints on the processing of syntactic ambiguity. Cognitive Psychology 24. 56–98.

Marcus, Gary F. 2004. The birth of the mind: how a tiny number of genes creates the complexities of human thought. New York: Basic Books.

Martin, Andrea E., and Brian McElree. 2008. A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. Journal of Memory and Language 58. 879–906.

——, and ——. 2009. Memory operations that support language comprehension: evidence from verb-phrase ellipsis. Journal of Experimental Psychology: Learning Memory & Cognition 35. 1231–9.

——, and ——. 2011. Direct-access retrieval during sentence comprehension: evidence from sluicing. Journal of Memory and Language 64. 327–43.

McElree, Brian. 1993. The locus of lexical preference effects in sentence comprehension: a time-course analysis. Journal of Memory and Language 32. 536–71.

——. 1996. Accessing short-term memory with semantic and phonological information: a time-course analysis. Memory & Cognition 24. 173–87.

——. 1998. Attended and non-attended states in working memory: accessing categorized structures. Journal of Memory and Language 38. 225–52.

——. 2000. Sentence comprehension is mediated by content-addressable memory structures. Journal of Psycholinguistic Research 29. 111–23.

——. 2001. Working memory and focal attention. Journal of Experimental Psychology: Learning, Memory, and Cognition 27. 817–35.

——. 2006. Accessing recent events. The psychology of learning and motivation, vol. 46, ed. by B. H. Ross, 155–200. San Diego: Academic Press.

——, and Barbara A. Dosher. 1989. Serial position and set size in short term memory: the time course of recognition. Journal of Experimental Psychology: General 118. 346–73.

——, and ——. 1993. Serial retrieval processes in the recovery of order information. Journal of Experimental Psychology: General 122. 291–315.

——, and Johanna Nordlie. 1999. Literal and figurative interpretations are computed in equal time. Psychonomic Bulletin & Review 6. 486–94.

——, Stephani Foraker, and Lisbeth Dyer. 2003. Memory structures that subserve sentence comprehension. Journal of Memory and Language 48. 67–91.

Miller, George, and Noam Chomsky. 1963. Finitary models of language users. Handbook of mathematical psychology, vol. 2, ed. by D. R. Luce, R. R. Bush and E. Galanter, 419–91. New York: John Wiley.

Murdock, Bennet B. Jr. 1971. A parallel-processing model for scanning. Perception and Psychophysics 10. 289–91.

Nairne, James S. 2002. The myth of the encoding-retrieval match. Memory 1. 389–95.

Oberauer, Klaus. 2002. Access to information in working memory: exploring the focus of attention. Journal of Experimental Psychology: Learning, Memory, and Cognition 28. 411–21.

——. 2006. Is the focus of attention in working memory expanded through practice? Journal of Experimental Psychology: Learning, Memory, and Cognition 32. 197–214.

Öztekin, Ilke, and Brian McElree. 2007. Proactive interference slows recognition by eliminating fast assessments of familiarity. Journal of Memory and Language 57. 126–49.

——, ——, Bernhard P. Staresina, and Lila Davachi. 2008. Working memory retrieval: contributions of the left prefrontal cortex, the left posterior parietal cortex, and the hippocampus. Journal of Cognitive Neuroscience 21. 581–93.

——, Lila Davachi, and Brian McElree. 2010. Are representations in working memory distinct from those in long-term memory? Neural evidence in support of a single store. Psychological Science 21. 1123–33.

Raaijmakers, Jeroen G. W., and Richard M. Shiffrin. 1981. Search of associative memory. Psychological Review 88. 93–134.

Ratcliff, Roger. 1978. A theory of memory retrieval. Psychological Review 85. 59–108.

——, and Gail McKoon. 1994. Retrieving information from memory: spreading-activation theories versus compound-cue theories. Psychological Review 101. 177–84.

Reed, Adam V. 1973. Speed-accuracy trade-off in recognition memory. Science 181. 574–6.

——. 1976. List length and the time course of recognition in immediate memory. Memory & Cognition 4. 16–30.

Sternberg, Saul. 1966. High speed scanning in human memory. Science 153. 652–4.

——. 1969. The discovery of processing stages: extensions of Donders' method. Acta Psychologica 30. 276–315.

——. 1975. Memory scanning: new findings and current controversies. Quarterly Journal of Experimental Psychology 27. 1–32.

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

Stowe, Laurie A. 1986. Parsing WH-constructions: evidence for on-line gap location. Language and Cognitive Processes 1. 227–45.

Tabor, Whitney, and Sean Hutchins. 2004. Evidence for self-organized sentence processing: digging in effects. Journal of Experimental Psychology: Learning, Memory and Cognition 30. 431–50.

Theios, John. 1973. Reaction time measurement in the study of memory processes: theory and data. The psychology of learning and motivation, vol. 7, ed. by G. H. Bower, 44–85. New York: Academic Press.

Townsend, James T., and F. Gregory Ashby. 1983. The stochastic modeling of elementary psychological processes. New York: Cambridge University Press.

Treisman, Michel, and Estelle Doctor. 1987. Memory scanning: a comparison of the dynamic stack and exhaustive serial scan models with an extension of the latter. Acta Psychologica 64. 39–92.

Van Dyke, Julie A. 2002. Retrieval effects in sentence parsing and interpretation. Unpublished PhD Dissertation, University of Pittsburgh, Pittsburgh, PA.

——. 2007. Interference effects from grammatically unavailable constituents during sentence processing. Journal of Experimental Psychology: Learning, Memory and Cognition 33. 407–30.

——, and Brian McElree. 2006. Retrieval interference in sentence comprehension. Journal of Memory and Language 55. 157–66.

——, and ——. 2007. Similarity-based proactive and retroactive interference reduces quality of linguistic representations. Annual CUNY Sentence Processing Conference, La Jolla, CA.

——, and Richard L. Lewis. 2003. Distinguishing effects of structure and decay on attachment and repair: a retrieval interference theory of recovery from misanalyzed ambiguities. Journal of Memory and Language 49. 285–316.

Verhaegen, Paul, John Cerella, and Chandramallika Basak. 2004. A working memory workout: how to expand the focus of serial attention from one to four items in 10 hours or less. Journal of Experimental Psychology: Learning, Memory, and Cognition 30. 1322–37.

Vosse, Theo, and Gerard Kempen. 2000. Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and a lexicalist grammar. Cognition 75. 105–43.

Wagers, Matthew W. 2010. Anti-local contexts improve the overall speed of dependency completion. Talk presented at the CUNY Sentence Processing Conference, New York, NY, March 19–21.

——, and Brian McElree. forthcoming. The span of focal attention in language comprehension.

Wanner, Eric, and Michael Maratsos. 1978. An ATN approach to comprehension. Linguistic theory and psychological reality, ed. by M. Halle, J. Bresnan and G. Miller, 119–61. Cambridge: MIT Press.

Watkins, Olga C., and Michael Watkins. 1975. Buildup of proactive inhibition as a cue-overload effect. Journal of Experimental Psychology 104. 442–52.

Wickelgren, Wayne A. 1977. Speed-accuracy tradeoff and information processing dynamics. Acta Psychologica 41. 67–85.

——, and Albert T. Corbett. 1977. Associate interference and retrieval dynamics in yes-no recall and recognition. Journal of Experimental Psychology: Human Learning and Memory 3. 189–202.

——, ——, and Barbara A. Dosher. 1980. Priming and retrieval from short-term memory: a speed-accuracy tradeoff analysis. Journal of Verbal Learning and Verbal Behavior 19. 387–404.