

# UC Santa Cruz

## UC Santa Cruz Previously Published Works

**Title**

Memory mechanisms for wh-dependency formation and their implications for islandhood

**Permalink**

<https://escholarship.org/uc/item/21v073q7>

**ISBN**

9781107008700

**Author**

Wagers, Matthew W

**Publication Date**

2013-10-17

**DOI**

10.1017/cbo9781139035309.009

Peer reviewed

Memory mechanisms for *wh*-dependency formation and their implications for islandhood

Matthew W. Wagers  
Department of Linguistics  
University of California, Santa Cruz

Mail correspondence to:

Matthew W. Wagers  
Stevenson College  
1156 High St  
Santa Cruz, CA 95064  
[mwagers@ucsc.edu](mailto:mwagers@ucsc.edu)

October, 2011, FINAL DRAFT, 51pp, 10350 words excluding references  
For J. Sprouse & N. Hornstein (Eds.), *Experimental Syntax and Island Effects*, Cambridge.

## 1. Introduction

The goal of this chapter is to provide a functionally-specified model of how long-distance dependencies are understood in real-time. In particular its emphasis will be how long-distance dependencies interact with working memory (WM) and how that interaction might inform our understanding of the form or origin of grammatical constraints. By long-distance dependencies, I refer primarily to unbounded displacement dependencies, such as those exhibited in *wh*-questions, relative clauses, topicalizations, comparative clauses, clefts, etc. The grammar appears to place no limit on the hierarchical distance that such dependencies can span. The head of the dependency -- the displaced constituent itself -- can occur many clauses away from the tail of the dependency -- a gap<sup>1</sup> or pronoun. The example in (1) illustrates this basic observation with relativization: the bold-face constituent, *cookies*, is only one clause distant from its gap in (1a). In (1b) and (1c), however, it is two and three clauses away, respectively.

- (1) a. The **cookies** [<sub>S</sub> that Phil bakes \_\_\_ ] contain pistachios.  
 b. The **cookies** [<sub>S</sub> that Billy bothers Phil to bake \_\_\_ ] contain pistachios.  
 c. The **cookies** [<sub>S</sub> that everyone knows that Billy bothers Phil to bake \_\_\_ ] contain pistachios.

There has long been a mutual influence between the study of long-distance dependencies and the study of WM constraints in language processing (Miller & Chomsky, 1963). Between its head and tail, a large number of processing events could occur which are

---

<sup>1</sup> The use of the terms *gap* or *gap site* is here intended to be neutral between grammars which posit an unpronounced constituent (a trace or copy; Brody, 1995, Chomsky, 1995, Frank, 2002, Stabler, 2000) and those which combine the displaced constituent with its subcategorizing host in other ways (e.g., Sag, Bender & Wasow, 2003, Bresnan, 2001, Steedman, 2000). In my view, psycholinguistic theories and data make no useful distinction between the alternatives at present (Phillips & Wagers, 2007).

essentially irrelevant to the long-distance dependency itself. At the syntactic level, these include events like establishing other thematic dependencies, interpreting modifier relations, checking case and agreement or resolving anaphora. The processing system must thus have a means of retaining information about the head of the long-distance dependencies until its ‘tail’ can be constructed, at which point it must be able to effectively recover the head. And it must do so without suspending the intervening processing events that span the dependency. These requirements describe exactly the sort of cognitive juggling act that working memory systems are proposed to accomplish (e.g., Baddeley, 1986, Miyake & Shah, 1999).

An important question to ask about how linguistic information interacts with WM is whether the grammar can generate structures that overwhelm WM capacity. For the moment, we will speak of capacity roughly as the amount of information that can be encoded and later recovered in a relatively loss-free fashion, as well as the extent of time for which that information can be maintained. Whether it is possible to give a general or useful characterization of WM capacity has driven considerable research in cognitive psychology for much of its modern history (Miller, 1956, Cowan, 2005). For language, we can imagine trivial examples that might overwhelm WM capacity by dint of their length. For example, I take it that no one can sensibly interpret a 256-clause sentence. The more interesting examples, however, feature syntactic or semantic complexities that seem to resist comprehension, even when the sentence is short. The prime example is the center self-embedded sentence:

(2) The cookies that the dog that Phil scolded tasted were burned.

Few speakers of English find (2) to be an acceptable sentence, even though it is straightforwardly generated by the language's phrase structure rules. However if we remove just one layer of embedding, the sentence becomes unremarkably acceptable:

(3) The cookies that the dog tasted were burned.

The extreme unacceptability of center self-embeddings is believed to stem from how the application or recognition of grammatical rules is constrained by processing capacity, though there are a diversity of proposals for the exact nature and locus of such a constraint (Ynvg, 1961, Frazier & Fodor, 1978, Stabler, 1994, Lewis, 1996, Gibson & Thomas, 1999, Vasishth, Suckow, Lewis & Kern, 2010, among many others).

We can raise an analogous question about *wh*-movement dependencies. A minor modification of sentence (1b) transforms it from an acceptable sentence to an unacceptable one:

(4) The cookies that Billy bothers Phil after he bakes \_\_\_ contain pistachios.

The unacceptability of a sentence like (4) has been standardly attributed in generative grammar to the violation of an island constraint. In this case, the dependency spans the boundary of an adjunct clause and this violates a condition on extraction (Huang, 1982; see **Chapter 0** in this volume). However, by analogy to center self-embedded sentences, we can ask whether some island-violating sentences might be freely generable by the grammar, yet unable to be processed during comprehension because of a capacity constraint. A number of proposals have related the unacceptability of particular island condition sentences to a confluence of factors, among which WM capacity figures strongly (Givón, 1979, Deane, 1991, Kluender & Kutas, 1993, Kluender, 2004, Hofmeister & Sag, 2010, among others; see

Chapters X, Y, Z in this volume). In order to evaluate these theories and their competitors, it is important to have a precise notion of how WM capacity is related to language comprehension and how long-distance dependencies do or do not strain it<sup>2</sup>.

In the next section, I will present a theory of working memory which has been emerging as a consensus among many memory theorists in the past decade (Nairne, 2002, McElree, 2006, Jonides et al. 2008) as well as some psycholinguists (Vasishth & Lewis, 2005). It has two interesting features: firstly, it does not assume there is any strong mechanistic discontinuity between memory in the short-term and memory in the long-term. The means of retention and retrieval are largely the same, whether they take place on the timescale of 500 milliseconds or 15 days. As a consequence, this theory of WM largely eschews specialized buffers and storage subsystems that were central in other theories (Baddeley, 1986). Secondly, the mechanisms for searching and retrieving information are optimized for the inherent features of stored representations and not the relations that hold between them (McElree, 2000, Van Dyke, 2007). This is, at first glance, problematic for language processing since constraints on grammatical dependencies are often characterized in both terms. For example, identifying the appropriate antecedent for a verb's reflexive argument is constrained both by independent properties of a potential antecedent ("feminine and singular and a noun phrase") and relational ones (e.g., "closest, c-commanding clause-mate"). In section 3, I will describe an empirically-grounded model for processing *wh*-dependencies which is compatible with the unitary model of memory. Finally, in section 4, I will attempt to identify whether any interactions in the model could

---

<sup>2</sup> Before doing so, it is important to stress that WM capacity is only one piece of the puzzle in understanding why certain sentences are easy to process and why certain sentences are difficult. It is, however, a necessary piece of the puzzle.

support an account of island constraints which reduces, in part, to working memory capacity or efficiency.

But let me first preview where the reader of this paper will end up. The question of how dependencies might strain the working memory system can be decomposed into two questions: what components of the dependency are required to be actively maintained, and what components must be recovered later via retrieval mechanisms. The argument developed in section 4, which addresses the problem of islandhood, takes the form of a 'threat assessment': how strong is the case that the WM system can be overwhelmed by dependency completion in the major island contexts? The threat, I believe, is too weak and too diffuse to heavily implicate a strain on the working memory system in the low acceptability of island-violating unbounded dependencies. I conclude that strong evidence is lacking that much information, if any, pertaining to the dependency is actively maintained across its span. Therefore, if islandhood depends on WM difficulty, the explanatory burden must rest with retrieval difficulty. There are several reasons to be skeptical, however, that the retrieval difficulty is severe enough to provide a useful explanation of islandhood. Some of these reasons are more conceptual in nature - such as analogy to the observation that retrieval pathways can be optimized when behavior is well practiced or routinized (Ericsson & Kintsch, 1995). Other reasons are more directly empirical - such as the fact that direct manipulations of retrieval difficulty have produced measurable but quite small effects on long-distance dependency completion (Van Dyke & McElree, 2006), or the fact that measures of WM efficiency show no correlation with the perceived severity of island violations (Sprouse, Wagers & Phillips, 2011).

Of course the validity of any threat assessment depends not only on the ‘known knowns’ but also the ‘known unknowns’ (and, more perniciously, the ‘unknown unknowns’; Rumsfeld, 2002). For this reason I try to highlight throughout where my assessment could be misguided, and in particular, what kinds of data are missing that might improve our understanding or shift the balance of explanation.

## **2. Working memory**

### *2.1 Introduction*

Miller’s famous 1956 paper discussed a variety of experiments testing the “span of immediate memory.” In these experiments (Hayes, 1952, Pollack, 1953), adult participants were challenged to recall lists of various symbols: binary digits, decimal digits, letters, and words. Regardless of symbol type, there was a remarkably narrow range of variation in performance: recall was generally quite accurate until list length exceeded 5-9 items (i.e., the “magic number”  $7 \pm 2$ ). This was striking to Miller because each of the symbol types conveyed different amounts of information: for example, a decimal digit conveys approximately 3.3 bits of information, while an English word conveys approximately 10 bits. Therefore he concluded that the capacity limitation on immediate memory was stated not in terms of information conveyed but in terms of a limited number of task-relevant encodings that could be successfully maintained and recalled in the short-term. Cowan (2005) has since argued that four is a more accurate estimate of the typical span of immediate memory for a variety of tasks.

The impact of Miller’s paper on linguistic research was immediate. Perhaps most notably, it motivated Yngve’s (1961) depth hypothesis, a proposal that the unacceptability



of doubly-embedded center self-embedding stemmed from a restriction on the depth of the parser's stack<sup>3</sup>. But the question immediately arises whether span sizes translate directly into an architectural notion, like number of buffers or slots in a stack. Here it is useful to draw a distinction between two concepts: 1) "working memory", broadly, as the sum of the many component parts and processes that support the encoding, retention and recollection of recently-encountered information; 2) "working memory", narrowly, as the particular mechanism for maintaining a specific piece information over the shortest term<sup>4</sup>. Inferring from a span number, be it 4 or 7, that there are a number of distinguished memory cells makes a commitment to the second concept. But it is logically possible that a relatively constant span derives from the first concept: that is, it derives from the interaction of many components and does not directly reflect the read-out of  $n$  working memory buffers, where  $n$  is the measure of span.

In recent years much evidence has accrued that supports the latter interpretation of span (see Conway, et al., 2008). In most cognitive tasks, the amount of information that can be concurrently maintained and made directly available to on-going processing is extremely limited (Broadbent, 1958, McElree & Doshier, 1989, 2006, Cowan, 1995, 2005, Jonides et al. 2008), smaller than the number of items that can be successfully recalled in a span task (McElree, 2006). This small amount of information is accessible to on-going cognitive processes with effectively very little delay. Information in this state is said to be in

---

<sup>3</sup> The depth hypothesis is a *grounded* account of a grammatical constraint (see, e.g., Fodor, 1978) since Yngve proposed to incorporate the depth-of-embedding limitation in the grammar as an adaptation to the memory limitation.

<sup>4</sup> This is sometimes called, more simply, short-term memory or STM. However, there seems to me to be enough confusion in psycholinguistics over the use of these terms that no terminological distinction will be introduced here.

the *focus of attention*<sup>5</sup>. Information outside of this state must be restored to the focus of attention to be useful, a process referred to very generally as retrieval (Anderson & Neely, 1996). Whether or not information needs to be retrieved is of considerable functional significance, since retrieved information takes more time to impact processing and retrieval is prone to error. There is consequently a functionally important interaction between the maintenance of the readily-accessible focal representations and the retrieval of other representations to displace or transform the current contents of focal attention.

In the next three sections we will attempt to untangle this interaction by answering the following questions:

- (5) **Q1** When must information be retrieved?
- Q2** How is information retrieved?
- Q3** What factors determine the success of retrieval?

### 2.2.1 *Question 1: When must information be retrieved?*

**Q1** amounts to the narrow capacity question: how much information can be concurrently maintained in focal attention. The greater the capacity of focal attention, the less often its current contents will have to be shunted to make way for new information. Therefore the capacity of focal attention contributes to the expected frequency with which memory retrieval operations occur. There is broad agreement that focal capacity is restricted. Specific estimates vary depending on task and stimulus structure (Cowan, 1995, 2001; Garavan, 1998; McElree, 2001, 2006; Oberauer, 2002). McElree and colleagues have argued that only one task-relevant representation is typically maintained in focal attention

---

<sup>5</sup> There is a strong analogy between a focal, non-focal split in information state and James's (1890) distinction between primary memory, evocatively called "the trailing edge of the conscious present", and secondary memory.

(McElree & Doshier, 1989, McElree, 2006, Wagers & McElree, 2011). Information not contained in that representation can only directly influence processing if it is retrieved to replace the focal representation. Their evidence comes from measuring the dynamics with which participants recognize or recall recently encountered data. This can best be illustrated with a concrete example from that research.

In a series of experiments, McElree (1996) asked participants to study 5-word lists, presented word by word. After the final word, a visual mask was first displayed followed by a test probe. Depending on the trial, participants judged either whether the test probe was in the memory list, whether it rhymed with an item in the list, or whether it was a synonym of an item in the list. Responses were collected at a variety of lags after presentation of the test probe, so that the point at which information begins accumulating and the rate at which that accumulation occurs could be measured. This approach, known as a response-signal method, measures the speed-accuracy trade-off (SAT) participants make in giving a response. It is more revealing than simply collecting reaction times (RTs), because RTs are a unidimensional measure of processing efficiency which conflate the speed at which cognitive processes run with the accuracy criteria that participants (implicitly) set in completing tasks (Wickelgren, 1976). What was consistent among all the trial types in McElree's experiment was that the most recently presented word achieved not only highest accuracy but it began its rise to accuracy the soonest. The remainder of the words in the list, regardless of their exact serial position, began their rise to asymptotic accuracy at the same later time, about 200 ms later. These data make the case that what matters for obtaining the speed advantage is that the test response depend on the last task-relevant

representation to occupy focal attention<sup>6</sup>. For the rhyme test, focal attention must include information about the word's phonology. For the synonym test, it must include information about the word's semantic features. Note that these results do not necessarily imply that all conceivable information about a word was available – upon presentation of the mask, participants were cued about which judgment was required so they could have transformed the last representation accordingly.

Finally, several findings indicate that the speed advantage is not uniquely linked to an item's being in final position. If experimental procedures are used which encourage participants to rehearse items from particular list positions immediately prior to test, then the focal advantage accrues to those items (McElree, 2006). Moreover, the structure of the word lists matter. If a word list can be parsed into multiple categories -- for example, names for furniture and names for flowers -- then the focal advantage accrues to the most recently encountered category, not the most recently encountered name (McElree, 1998). Finally, evidence for focal attention comes from data other than SAT studies, including RT distributional analyses (Oberauer, 2002, 2006) and fMRI studies of activation in hippocampus and inferior frontal gyrus (Öztekin et al., 2008; see Cabeza et al., 2003).

### 2.2.1 *How is information retrieved?*

**Q2** asks for the mechanism by which other previously-constructed representations are restored to the focus of attention. The answer to **Q2** depends in part on the architecture of memory: that is, what lies 'beyond' the focus of attention. McElree (2006) draws a distinction between bi-partite and tri-partite working memory architectures. In bi-partite

---

<sup>6</sup> Additionally, the fact that study and test were interrupted by a visual mask allow us to dismiss any low-level perceptual account of the advantage.

architectures, information can only occupy two states: *active*, in the focus of attention, or *passive*, in its durably-encoded long-term memory (LTM) state. In tri-partite architectures, an intermediate state exists for information not currently being processed which is nonetheless not simply in LTM. For some models, like Baddeley (1986), this intermediate state corresponds to a short-term memory store that is separate from long-term memory – like a buffer. Other models, like Cowan (1995, 2001) or Oberauer (2002), reject a separate short-term store but claim that a small amount of recently used information is so highly activated as to be functionally distinct from other LTM encodings.

The bi-partite architecture makes a clear prediction about the mechanisms of retrieval. Because encodings outside of the focus of attention are in the same state as LTM, then it is predicted that the retrieval of recently-encoded information will proceed in largely the same way as information encoded in LTM. Though it is not likely the case that there is a ‘single mechanism’ of retrieving from LTM, the dominant mechanism appears to be associative, or content-addressable, retrieval (Clark & Gronlund, 1996, Gillund & Shiffrin, 1984, Hinton, 1989, Hintzman, 1988, Kawamoto, 1988, Murdock, 1982, Plaut, 1987). Content-addressable retrieval refers to the use of the contents of memory encodings themselves in the access procedure. Inherent features of the desired encoding are used as probes to identify matching encodings. Thus content-addressable access may be contrasted with access to information that proceeds by iteratively inspecting storage locations, i.e. a search. Content-addressability is implementable in a variety of architectures and is characteristic of most contemporary memory models (see Clark & Gronlund, 1996, for a

review).<sup>7</sup> The key advantage of building content-addressability into a memory system is that, when information of a certain type is desired, it is often not necessary for the system to consult or otherwise be influenced by irrelevant information. This contrasts with search procedures which require comparisons of the desired information with each memory record in the search set. Content-addressable retrievals yield retrieval times that are independent of the size of the search set. Searches, in contrast, yield retrieval times that are directly proportional to set size. The second advantage of content-addressability is thus its speed.

### 2.2.3 *Question 3: What factors determine the success of retrieval?*

The major determinant of retrieval success is the match between the information used at retrieval – the cues – and the desired encoding. If the combination of cues used at retrieval is sufficiently distinct, then retrieval success will be high. However, if they apply to many different encodings in memory, then the wrong encoding may be retrieved. This phenomenon is referred to as similarity-based interference (Anderson & Neely, 1996). This is the tradeoff for the fast access times associated with content-addressable memories: irrelevant encodings that are similar to the desired encodings can negatively impact processing. This problem is potentially very acute for linguistic representations. Because linguistic representations are recursive, compositional objects built out of a relatively small repertoire of atomic parts, they contain highly self-similar subparts.

---

<sup>7</sup> The address labels in a random-access memory (RAM) are themselves arbitrary, but direct, content-addressable access to specific encodings can nonetheless be achieved using hash-coding (Kohonen, 1977). The crucial issue is thus not how the memory is physically implemented, but how many memory accesses are necessary to retrieve a desired encoding.

On the one hand, much available evidence suggests that content-addressability nonetheless prevails in memory retrieval for language processing. There is growing evidence of similarity based interference in certain environments (Gordon, Hendrick, & Johnson, 2001, Van Dyke & Lewis, 2003, Lewis & Vasishth, 2005). As an example of both similarity-based interference and its relevance to language processing, consider one of the experimental sentences from Van Dyke & Lewis (2003), in (6). In this sentence, the subject of the embedded clause, *the student*, should be paired with the predicate, *was standing*.

- (6) The secretary forgot that ...  
 the student who thought that the exam was important was standing in the hallway.
- 
- The diagram shows a dotted line with three upward-pointing triangles (↑) and one downward-pointing triangle (↓). The first triangle is under 'the student', the second is under 'was', and the third is under 'standing'. A vertical dashed line with a downward-pointing triangle is under 'the exam'.

However, the presence of a full lexical subject (*the exam*) in the intervening relative clause can impact the dependency formation process, rendering it more slow or less accurate. This finding has been replicated in a number of contexts (Van Dyke, 2007, Wagers, 2008), with the major determinant of retrieval success appearing to be whether or not a linearly intervening, grammatically inappropriate [Spec,TP] position is occupied. Consistent with this evidence, Lewis & Vasishth (2005) and Lewis, Vasishth & Van Dyke (2006), on the basis of their ACT-R model of sentence processing, have argued that similarity-based retrieval interference is a significant determinant of comprehension success. In ACT-R, or in any model that incorporates content-addressability, the fact that the encodings of the grammatically-appropriate subject in (6) and the grammatically inaccessible subject overlap in some of their features is what renders them liable to retrieval interference. Many instances of grammatically-inaccurate performance seem to yield nicely to a retrieval-based account, such as patterns of case and agreement attraction (Wagers, Lau & Phillips, 2009).

On the other hand, there is a large body evidence indicating that much of sentence processing is grammatically accurate, and interference-robust (Phillips, Wagers, Lau, 2011). An important research question is how these two sets of empirical observations might be reconciled. It may be that different linguistic phenomena are processed by distinct memory mechanisms. According to this view, fast, interference-prone memory operations could characterize some kinds of dependencies in language; while slower, search-style operations characterize others. A second (non-exclusive) possibility is that the systems which encode linguistic representations and manage the cues used at retrieval have become effectively adapted to the nature of linguistic representations, such that optimally diagnostic cue sets are used in retrieval. Ericsson & Kintsch (1995), in their theory of long-term working memory, have made essentially this proposal to account for expert performance in well-practiced, narrow domains -- for example, chess. A skilled player can accurately recall complex (legal) chess positions from her recent games not because she has an exceptional memory. Instead, it is argued she has a well-developed skill at encoding specific episodes of chess in a way that will make them selectively targetable with an effective and frugal retrieval structure, despite the fact that there are many abstract similarities from chess game to chess game. Ericsson & Kintsch write, “ ... the acquired memory skill involves the development of encodings for which the subject can provide controlled access to significant aspects of the encoding context and thus indirectly to the desired information in a manner consistent with the encoding-specificity principle (Tulving, 1983)” (1995, p. 10). It seems likely the same skills are developed in language comprehension. Though the component memory processes are themselves fast, limited and error-prone, comprehenders may strategically encode each new dependency to guard against similarity-based interference.



In Section 4, I will illustrate one possible strategy for doing so with long-distance dependencies.

In this section, I have sketched an account of the memory architecture which forms one component of the language processing system. To summarize: this architecture allows for minimal concurrent activation of cognitive representations; it is therefore frequently cycling representations between the active and passive states, guided by content-addressable retrieval. By virtue of its content-addressability, cognitive operations are liable to fail when similar representations compete at retrieval. In the next section I shall turn to what the parsing of *wh*-dependencies looks like in such an architecture.

### 3. The time course of long-distance dependency formation

#### 3.1 *The basic generalization: island sensitive, active dependency formation*

The last 30 years of psycholinguistic research have set several important empirical boundaries on any theory of *wh*-processing. The first observation is that *wh*-processing is a highly incremental, predictive process: a property I'll refer to as *active dependency formation*. An important cue to a long-distance dependency's interpretation is the absence of a particular constituent, i.e., the gap. The parser must 'detect' the gap -- that is, identify its subcategorizing syntactic head -- in order to correctly thematically integrate the filler. Jackendoff & Culicover (1971), in discussing patterns of acceptability in dative questions, proposed that gaps were only hypothesized as a last resort: only when an obligatory constituent was absent would the gap be 'detected.' The Augmented Transition Network model of Wanner & Maratsos (1978) embodied this strategy by only analyzing and attaching displaced NPs if it could not otherwise recognize a legal VP constituent. Fodor (1978), however, argued that the last resort strategy was too strong. For example, it predicted comprehension difficulty that does not exist with extraction from optional transitives. The sentence in (7) illustrates this observation with the verb *read*: there is no intuitive difficulty in recognizing the filler-gap dependency, yet "read again to her son" is a legal VP even without a gap.

(7) Which book did the tired mother have to read again to her son?

Fodor proposed that gaps should be able to be postulated in advance of direct evidence for their location. In a case like (7), Fodor argues, readers preferentially posit a direct object gap because of the likelihood of *read* to take a direct object.

Direct experimental evidence has broadly supported the idea that the parser completes long-distance dependencies without waiting for unambiguous evidence of the gap position: that is, it is an active dependency completion parser. Phillips & Wagers (2007) presents a review of the evidence that supports this conclusion, evidence which spans a diverse array of experimental methodologies as well as one of the broadest cross-linguistic samples I know of in psycholinguistics<sup>8</sup>. Here I will mention only a few experiments. One of the earliest demonstrations comes from the *filled-gap effect* (Crain & Fodor, 1985, Stowe, 1986). Stowe (1986) compared self-paced reading times for sentences containing a displaced *wh*-phrase, like (8a), with matched sentences without a *wh*-dependency, like (8b):

- (8) (a) My brother wanted to know **who** Ruth will bring *us* home to \_\_\_ at Christmas.  
 (b) My brother wanted to know if Ruth will bring *us* home to Mom at Christmas.

The direct object NP, *us*, led to longer-reading times in (8a) compared to the same NP in (8b). Stowe construed this contrast as a surprise effect, which derived from an initial direct object interpretation of the filler. Using the same logic, Lee (2004) has shown that prior to its direct object interpretation, the filler is interpreted as originating in subject position.

---

<sup>8</sup> CROSS-LINGUISTICALLY: Dutch (Frazier 1987, Frazier & Flores D'Arcais 1989, Kaan 1997), German (Schlesewsky, Fanselow, Kliegl & Krell 2000), Hungarian (Radó 1999), Italian (de Vincenzi 1991), Japanese (Aoshima, Phillips & Weinberg 2004), Russian (Sekerina 2003).

CROSS-METHODOLOGICALLY: Electrophysiology using EEG (Garnsey et al. 1989, Kaan, Harris, Gibson & Holcomb 2000, Phillips, Kazanina & Abada 2005) and MEG (Lau, Yeung, Hashimoto, Braun & Phillips 2006), the 'stops making sense' task (Tanenhaus, Stowe & Carlson 1985, Boland et al. 1995), eye-tracking (Traxler & Pickering 1996), crossmodal lexical priming (Nicol & Swinney 1989, Nicol, Fodor & Swinney 1994), anticipatory eye movements (Sussman & Sedivy 2003).

Electrophysiological studies have provided convergent evidence. For example, Garnsey et al. (1989) varied the plausibility of filler-verb combinations, as in (9a-b), and observed immediate detection of the semantic anomaly on the verb which hosts the gap, as indexed by an N400 evoked response.

- (9) (a) The businessman knew **which customer** the secretary called \_\_\_ at home.  
 (b) The businessman knew **which article** the secretary called \_\_\_ at home.

Traxler & Pickering (1996) showed that reading times measured in eye-tracking experiments increase under a similar manipulation.

There are a number of possible (non-exclusive) motivations for active dependency formation. As Fodor (1978) argued, a non-active parser -- one which is cued by failure to detect an obligatory constituent -- will often unknowingly make mistakes. An active parser might often make mistakes -- which is the very premise of the filled-gap effect -- but it will nonetheless obtain a clear error signal. And when it does not make mistakes, the active parser will be able to establish some crucial aspects about the sentence's interpretation sooner rather than later. Wagers & Phillips (2009) argued that an important and distinct motivation for active dependency formation derives from a pressure to satisfy open grammatical licensing requirements as soon as possible, a view which can be identified with principle-based parsing (Pritchett, 1992, Weinberg, 1992). Using a plausibility manipulation, they contrasted the processing of across-the-board raising extractions from conjoined VPs (as in 10a-b) with potential parasitic gap hosts (as in 11a-b). While extraction is essentially obligatory from both conjuncts of conjoined VPs, it is optional in the case of parasitic gaps. Wagers & Phillips found evidence for active dependency completion for the second gap in the obligatory extractions (10), but not for the potential, optional extractions (11).

- (10) *Coordinated VP, Plausible*  
 (a) The **wines** which the gourmets were energetically discussing \_\_  
 or slowly *sipping* \_\_ during the banquet were rare imports from Italy.  
*Coordinated VP, Implausible*  
 (b) The **cheeses** which the gourmets were energetically discussing \_\_  
 or slowly *sipping* \_\_ during the banquet were rare imports from Italy.
- (11) *Potential parasitic gap, Plausible*  
 (a) The **wines** which the gourmets were energetically discussing \_\_  
 before slowly *sipping* the samples during the banquet were rare imports ...  
*Potential parasitic gap, Implausible*  
 (b) The **cheeses** which the gourmets were energetically discussing \_\_  
 before slowly *sipping* some wine during the banquet were rare imports ...

The parser thus appears to be sensitive to what the grammar requires of well-formed long-distance dependencies in a way that affects whether or not the dependency is completed actively.

This finding leads us to a broader generalization, which is the second -- and for this chapter, more interesting -- property of long-distance dependencies completion: *island sensitivity*. Though the parser completes dependencies actively, it does not do so at all costs. The across-the-board extraction facts explored by Wagers & Phillips (2009) strengthens a much larger body of observations that island domains are respected in online processing. Across many studies the parser overwhelmingly does not posit gaps inside island domains, evidence that Phillips (2006) reviews. Traxler & Pickering (1996), an eye-tracking study, provides an illustrative example of this fact for relative clause islands. Firstly they consider a plausibility contrast in a simple relativization:

- (12) We like **the book / the city** that the author *wrote* unceasingly and with great dedication about \_\_ while waiting for a contract.

In case of either filler, the ultimately correct analysis is unexceptionable: books and cities are both plausible things to write about. In the initial analysis, however, the active parser

interprets the filler as the direct object of *wrote*. While it is plausible to write books, it seems impossible to ‘write cities.’ Consistent with similar experiments, like Garnsey et al. (1989), Traxler & Pickering observe a slow-down at the verb *wrote* for the implausible filler compared to the plausible one. In a second contrast, they subordinated *wrote* inside a relative clause, as in (13), thus making any potential gap hosted by *wrote* inaccessible to the filler in the matrix clause.

(13) We like **the book / the city** that the author [<sub>RC</sub> who *wrote* unceasingly and with great dedication ] saw \_\_\_ while waiting for a contract.

In this case, no effect of filler plausibility is observed at the lure verb *wrote*. The authors conclude that comprehenders never entertain the island-violating dependency as an analysis.

One straightforward interpretation of the island sensitivity findings is that the grammatical restriction against long-distance dependencies terminating inside certain domains causes the parser to refrain from positing a dependency inside of them. However a potential worry can be raised about such findings: because island sensitivity is reflected by the absence of any effect (with the exception of Wagers & Phillips, 2009), it may be that the parser does not specifically refrain from positing dependencies, but that it is simply unable to do so. In other words, demonstration of island sensitivity in long-distance dependencies completion experiment is also potentially evidence for the proposal that islands are real-time epiphenomena (Deane, 1991, Pritchett, 1991, Kluender & Kutas, 1993). Phillips (2006) attempted to respond to this proposal for subject islands (Pritchett, 1991, Kluender, 2004). Extraction from subjects normally appears to be ill-formed (14).

(14) \* What did the attempt to repair \_\_\_ ultimately damage the car?

However, as with VP adjuncts, parasitic gaps are legal in certain subjects, particularly those whose head takes an infinitival complement, as in (15).

(15) What did the attempt to repair \_\_\_<sub>PG</sub> ultimately damage \_\_\_ ?

Phillips (2006) shows that comprehenders posit gaps inside just those subject phrases which can support a parasitic gap. Thus, he argues, it cannot be the inability of comprehenders to create a dependency in a subject environment that accounts for the ill-formedness of sentences like (14)<sup>9</sup>.

### 3.2 *Integration with the working memory architecture*

With these two basic facts about long-distance dependency completion in hand -- dependency completion is active but it is island-sensitive -- we can attempt to integrate the theory of memory explored in section 2 with the time-course facts. There are two essential questions to answer:

(16) *Maintenance*

Is any information actively maintained in memory while a long-distance dependency remains unresolved?

(17) *Retrieval success*

---

<sup>9</sup> Phillips (2006) and Wagers & Phillips (2009) present an interesting contrast: comprehenders are willing to undertake a subject parasitic gap analysis, but not an adjunct parasitic gap analysis. Wagers & Phillips (2009) argue that the relative linear ordering of the parasitic and licensing gap may explain this difference: comprehenders are willing to undertake the parasitic gap analysis while they are still searching for the licensing gap, but they will not actively undertake a parasitic gap analysis if the licensing gap is already resolved. That claim remains to be directly tested.

What information is used to cue the retrieval of the filler, once the gap is postulated?

(16), the maintenance question, has often suggested itself as a source of difficulty in comprehension (Gibson, 1998, 2000, Fiebach et al., 2002, Kluender & Kutas, 1993). To use a phrase sometimes encountered in discussions of this phenomenon, the trouble with long-distance dependencies may be that they require a filler to be ‘carried’ forward in time. (17), the retrieval success question, has figured somewhat less prominently in discussion of long-distance dependencies comprehension but it is closely related to the view that fillers have to be reactivated. In some recent research, Hofmeister & colleagues (2010) have fingered retrieval success as a major determinant of complexity in long-distance dependency formation.

I will first attempt to answer the maintenance question. While the answer here is unfortunately probably the least clear, the data suggest that, overall, only very little is actively maintained of the filler’s contents. If this is the correct conclusion, it is convergent with the focus of attention limitation on concurrent maintenance introduced in Section 2.2.1. It also suggests that the hypothetical cost of carrying filler information forward in time is itself not a major contributor to complexity to be reckoned with in prospective theories of islandhood.

### 3.2.1 *Maintenance*



Wanner & Maratsos (1978) was an early proponent of the idea that there is a storage cost for the filler in an incomplete long-distance dependency. In their ATN model, the cost stemmed from the fact that the filler was kept in a distinguished memory register, called the HOLD cell<sup>10</sup>, until it was analyzed at its gap location. The Active Filler Strategy, developed in various forms in Frazier (1987), Frazier & Flores D'Arcais (1989) and Frazier & Clifton (1989), can be seen as inspired by the spirit of the HOLD cell hypothesis, though it is not explicitly committed to maintenance. The Active Filler Strategy states that the parser should prefer attaching gaps to full NPs while a long-distance dependency is being resolved. However, there is a more interesting dynamic underpinning to the Active Filler Strategy than merely enforcing a preference: if the filler could somehow effectively outcompete the bottom-up input for attachment at potential argument sites, then the observed 'preference' would follow as a consequence of this competition. It would, of course, be necessary to have a mechanism to promote this competition, to covertly enter the filler into the parser's workspace. In Frazier's terms, the filler had to be kept non-'inert'.

Consistent with a maintenance hypothesis, long-distance dependencies do robustly exact a cost on linguistic performance, as measured by almost any method (e.g., Wanner & Maratsos, 1978, King & Just, 1991, Sprouse, Wagers & Phillips, 2011). However, Gibson & Grodner (2005), among others, have tied this cost to the re-integration of the filler at the gap site, and not the openness of the dependency per se. Of course it is possible that the reading time evidence supporting that conclusion may simply not be sensitive to memory

---

<sup>10</sup> More specifically, the proposed cost stemmed from the fact that the filler phrase was not stored as an analyzed constituent, but only as a string-to-be-analyzed. The assumption that the filler was unanalyzed was a consequence, in part, of the architecture of augmented transition networks; but linking the cost to its string encoding seemed to fit nicely with an early finding in psycholinguistics that imposing a syntactic analysis significantly improved the effective span of immediate free recall (see Fodor, Bever & Garrett, 1974).

load in the right way. Indeed researchers working with electrophysiological methods have occasionally reached a seemingly opposite conclusion. EEGs obtained while participants read object-extracted filler-gap dependencies or long-distance questions show a sustained anterior negativity (SAN) while the dependency remains open (King & Kutas, 1995, Fiebach, Schlesewsky & Friederici, 2002, Phillips, Kazanina, & Abada, 2005). The SAN has previously been implicated in explicit memory load tasks (Ruchkin, Johnson, Canoune & Ritter, 1990) and thus its presence in open filler-gap dependencies has been interpreted as a rather direct reflection of the memory load consumed by actively maintaining the filler.

This interpretation is qualified by two points. Firstly, the SAN does not reflect a cumulative effect that accrues or is renewed at each word, but instead it derives mainly from the first words of the dependency (King & Kutas, 1995, Phillips, Kazanina, & Abada, 2005). More importantly, the SAN studies do not speak to what the actual contents of the filler are when the dependency is unsatisfied. In discussing the SAN, Fiebach et al. (2002) are careful to point out that the electrophysiological effect itself does not discriminate between alternative accounts of what is being maintained. It could be a full semantic or syntactic representation of the filler, or perhaps a subset of the features of those representations. Alternatively, it may not contain the filler's content at all but instead register the existence of an (unsatisfied) prediction for a syntactic environment that allows completion of the dependency, as in Dependency Locality Theory (Gibson 2000).

More direct evidence about the maintained contents of the fillers come from cross-modal lexical activation studies and probe recognition tasks (Bever & McElree, 1988, Nicol & Swinney 1989; Nicol, Fodor & Swinney 1994). In cross-modal lexical decision, auditory word-by-word sentence comprehension is interrupted unpredictably by an on-screen

lexical decision task. For example, a test sentence containing a long-distance dependency might be given as follows:

(18) The doctor that the visitor from the prestigious University met in the lobby seemed to be in a hurry.

The lexical decision task would then probe with either a semantic associate of the filler *doctor* -- for example, 'nurse' -- or a suitable control word -- for example, 'mechanic.' By arranging to interrupt sentence comprehension at enough points between the filler and the gap (over the course of the experiment), it is possible to derive a temporal profile of the filler's activation in memory. In all such studies, there is evidence that semantic features of the filler are not maintained and must be reactivated later into a state suitable for integration. In particular the filler's semantic associates are primed immediately following the introduction of the filler in the sentence (Nicol & Swinney 1989; Nicol, Fodor & Swinney 1994). However the priming effect declines sharply during the rest of the sentence and only rises again at the gap site. McElree (2001) documented a similar pattern using a probe recognition task<sup>11</sup>. Thus the reactivation studies suggest that the filler is not concurrently

---

<sup>11</sup> In this task, participants had to report whether the interrupting word was a synonym of an adjective contained in the filler phrase. For example, if the filler were 'the brave puppy', a related probe word might be 'courageous.' This design feature avoids a contention raised by McKoon & Ratcliff (1994) about earlier studies. There they argued that the RT facilitation taken to be evidence for reactivation could reflect the fact that interruption points near the gap were also good candidate integration sites for the lexical decision target. Targets happened to be nouns and the intervening interruption points were less good fits to introduce a noun than the direct object position. According to this interpretation of the data, the cross-modal lexical decision task does not reflect activation state. The McElree (2001) study partially blunts this criticism by probing with a word which could not syntactically or semantically be integrated at the point of interruption.

maintained in a rich enough form to prime its lexical associates or otherwise facilitate processing of related words.

Finally, Wagers & Phillips (2009) reported that the slow-down effect which is normally robustly observed at the verb when the filler is implausible can be seriously attenuated if the distance between filler and verb is increased only slightly. When a relative clause long-distance dependency was serially lengthened by attaching a 5-word prepositional phrase to the intervening subject, the slow-down effect no longer surfaced on the verb and instead moved to a post-gap position. Wagers & Phillips (2009) proposed that, for longer long-distance dependencies, the anomaly detection effect could be greatly delayed because semantic integration might no longer take place immediately at the verb, but only after the semantic details of the filler could be adequately retrieved<sup>12</sup>. Wagers & Phillips (2011) replicated their original finding for the plausibility effect but reported that at dependency lengths comparable or longer, the filled-gap effect was preserved. The modified filled-gap paradigm they employed contrasted DP and PP extraction (as in Lee, 2004). The fact that this index of active dependency completion survived multiple dependency lengths suggests that, unlike fine semantic details, at least coarse-grained

---

<sup>12</sup> In light of the generalizations developed in section 3.1, this finding might seem at first surprising. It is useful to note, however, that past studies of filler-gap dependency construction have been heavily skewed to very short distances. In a sample of 21 influential experiments performed between 1986 to 2004, Wagers (2008) found there were on average only 2.9 linear interveners, or 1.8 constituent interveners, between filler and gap. These experiments spanned different paradigms: 10 self-paced reading studies, 5 eye-tracking studies, 3 sensicality monitoring studies, 3 cross-modal priming studies. Tanenhaus, Stowe & Carlson (1985); Stowe (1986); Swinney et al. (1988); Frazier & Clifton (1989); Boland et al. (1995); Pickering, Barton & Shillcock (1994); Pickering & Traxler (1996); Traxler & Pickering (1996); Clahsen & Featherston (1999); McElree (2000); Sussmann & Sedivy (2003); Aoshima, Phillips & Weinberg (2004); Conklin, Koenig & Mauener (2004); Lee (2004).

syntactic category information about the filler may be maintained (consistent also with Gibson 2000).

### 3.2.2. *Retrieval*

If the filler's contents are not maintained in a privileged state, they must be retrieved once a suitable gap host is identified in the input or constructed based on expectations of upcoming input. There are several possibilities for how the correct filler constituent is then retrieved. Lexical information, such as the syntactic or semantic restrictions the verb places on its arguments, may be one source of retrieval cues for the filler. Van Dyke & McElree (2006) provided some direct evidence that the filler is retrieved at the site of integration based on verb properties, and that this retrieval is subject to interference. They examined the processing of clefts under a memory load manipulation. In half the experimental conditions, participants were presented with a list of three nouns at the start of the trial, which would have to be recalled after the sentence comprehension task (Load conditions).

For example:

(19) **table - sink - truck**

Participants then read sentences like the following:

(20) It was **the boat** that the guy who lived by the sea *sailed / fixed* \_\_\_ in two sunny days.

Two possible critical verb types could occur in the sentence. In half the conditions, exemplified by 'sailed', the critical verb was not a good fit for the memory load nouns and thus generated low interference: it is not plausible to sail a table, sink or truck. In the other

half of the conditions, the critical verb was a good fit and thus generated high interference: tables, sinks and trucks are, like boats, fixable things. In the critical verb region they reported an interaction of interference and load conditions. Reading times between the two verb types were identical when there was no memory list; however, under Load conditions, high interference verbs were read more slowly. Thus the goodness of the match between the verb and the filler as its object determined the ease of comprehension in this experiment.

The plausibility effect studies, like Garnsey et al. (1989) and Traxler & Pickering (1996), provide an interesting constraint on the retrieval process. Recall that in those studies, an anomalous verb-filler combination was detected very early at the verb. This suggests that even if a filler is not a good object for the verb, it may nonetheless be retrieved. However, previous studies have not always carefully distinguished between purely selectional restrictions and violations of real-world expectations, so some caution is required. For example, in Traxler & Pickering (1996), the verb-filler combination 'shoot the garage' is certainly unusual, but it does not obviously violate a selectional restriction. Boland et al. (1995) found that active dependency formation was not pursued for anomalous verb-object combinations if the verb was a verb like 'remind' -- which combines with an animate object and a clause controlled by that object -- and the filler was an inanimate noun like 'movie.' Pickering & Traxler (2001) obtained convergent evidence for this finding and argued that certain simple selectional features, particularly animacy, could be used to filter dependency completion. One mechanism for achieving this filtering is to block retrieval of the filler if it is known not to match the gap host's selectional restrictions or if an initial retrieval based on a positive selectional requirement fails.

Subcategorization information may also be well suited as a cue. For example, intransitive verbs like ‘talk’ or ‘arrive’ do not combine with a direct object, so one can ask what kind of gaps positions are actively projected in VPs headed by intransitive verbs<sup>13</sup>. The subcategorization frame for such verbs, containing no second argument, may prevent -- or perhaps only make less effective -- the retrieval of the filler. The evidence is currently somewhat mixed. If we pay attention to two quite recent studies, we find Staub (2007) on the one hand, finding no evidence for active dependency formation at intransitive verbs, and Omaki, Lau, Davidson White & Phillips (2011) on the other, finding positive evidence for active dependency formation. The empirical landscape is somewhat clearer with regards to relative subcategorization frequency. Both Pickering & Traxler (2003) and Frazier & Clifton (1989) find that, when DP and PP subcategorization frames compete, the filler is always initially analyzed as being extracted from DP position, regardless of preference (but c.f. Stowe et al. 1991, and discussion in Fodor, 1978).

Finally, although the gap host can provide a rich set of cues with which to retrieve the filler, we also know that it cannot be the only source of information. Verb-final languages like Japanese show evidence of active long-distance dependency formation well in advance of the verb (Aoshima et al., 2004, Nakano et al., 2002). If Omaki, Lau, Davidson White & Phillips (2011) are correct that the parser sometimes constructs gaps hosted by intransitive verbs, then even English may be a language in which long-distance dependency formation occurs without verb information. Nothing should seem outlandish in that

---

<sup>13</sup> The necessary distinction may need to be finer than simply whether the verb takes a second argument. Unergative verbs, like *run*, which only take an external argument, may differ from unaccusative verbs, like *arrive*, whose single argument may originate within the VP. Likewise, the possibility of diathesis phenomena such as causative/inchoative and unergative/transitive alternations could affect the retrieval; in the psycholinguistic literature these alternations have usually been addressed only indirectly, under the cover term of a particular verb’s ‘optional intransitivity’.

proposal, as long as it is allowed that the construction of a particular syntactic phrase is not strictly dependent on its head having been pronounced. Interestingly Nakano et al. (2002) provide evidence from cross-modal lexical priming for pre-verbal activation in the argument field of a gapped Japanese VP. Moreover, they find evidence that reactivation is strongest among individuals who have high WM scores. These two pieces of evidence suggest that pre-verbal dependency formation may actually involve retrieval of the displaced argument, and not merely formation of a content-free syntactic representation. If there is retrieval from outside of the focus of attention, then the cues must derive from parsing rules and expectations and not from information contained in a specific lexeme.

#### **4. Implications for island constraints**

##### *4.1 Recap*

If one attempts to put together the theoretical and empirical pieces from sections 2 and 3, the following sort of picture emerges. When a displaced phrase is encountered during language comprehension, the parser actively seeks to integrate it with its gap host. However, other kinds of processing events must occur in the meanwhile. Because of the strong narrow capacity limitation on working memory, all or almost all of the displaced phrase's encoding must be displaced from the focus of attention. There is thus relatively little maintenance possible during the processing of a filler-gap dependency and the evidence suggests that there is very little maintenance required. This does not mean that no filler information is preserved in advance of retrieval. The proposal, by Pickering & Traxler (2001), that some kinds of selectional requirements set early filters on dependency formation suggests that it could be useful to preserve a bare-bones feature set for the filler. Furthermore the evidence from Wagers & Phillips (2011) that the filled-gap effect survives



long dependency lengths suggests at least coarse syntactic category information may be preserved. Stronger evidence is needed to substantiate these claims. The kind of focus of attention studies described in Section 2 would provide the most convincing evidence.

The parser cues retrieval at some point in advance of the gap site. Although it is typically taken to be the verb that cues retrieval in a verb medial language like English, the evidence is scant that no dependency formation occurs in advance of the verb. The cue set or retrieval structure used during this process can come not only from (1) a syntactic head, like the verb, but also (2) the internal rules of the parser; (3) any information maintained in the focus of attention; and potentially (4) non-linguistic information from the general context. The best case scenario is one in which the cue set compiled is sufficiently specific to activate one encoding. In general the parser can improve the likelihood of successful retrieval by ensuring that the information from multiple (uncorrelated) sources contributes to the retrieval set. As mentioned in Section 2, experts in particular domains, like chess or music, develop large and accurate memories in their domain of expertise by learning how to adaptively assemble retrieval structures. It seems reasonable, though as yet undemonstrated, to expect that language comprehenders do the same.

In the accompanying figure, I lay out one strategy comprehenders could use to assemble their retrieval structures in a way that would strengthen the likelihood of a successful grammatical outcome. The basic idea is that comprehenders should maintain a very sparse feature set from the original filler: indeed, it is less important to retain grammatical features, and more important to retain features that are specific to the event of encoding the filler, something like a randomly generated key or a hash (see Polyn, Norman & Kahana, 2009, for a formalization of this idea, and Tulving & Thomson, 1973, for

theoretical grounding). If comprehenders combine the maintained features with the retrieval cues generated at the retrieval site, then retrieval inference would be significantly dampened<sup>14</sup>.

---

<sup>14</sup> As a reviewer points out, it would be important to establish whether sparsification of the original filler representation would solve the problem set by the limited maintenance capacity of the system. This is clearly an empirical question. One can imagine, following Kluender & Kutas (1993), that encountering another A-bar operator, as in *wh*-islands or relative clause islands, leads to displacement of even the sparse representation of the filler. If that were true, then it would be important that retrieval still be able occur when this information is unavailable, and that it can generally succeed. As I discuss below, retrieval interference effects themselves vary in severity, so it is plausible that retrieval outcomes would be successful without the 'robustness' provided by carrying forward some episodic information. Note that this is likely true since the mere existence of an island domain in the midst of an unbounded dependency *that does not terminate in that domain* does not seem to affect the resolution of that dependency. See Wagers (2008), Experiments 9-10.

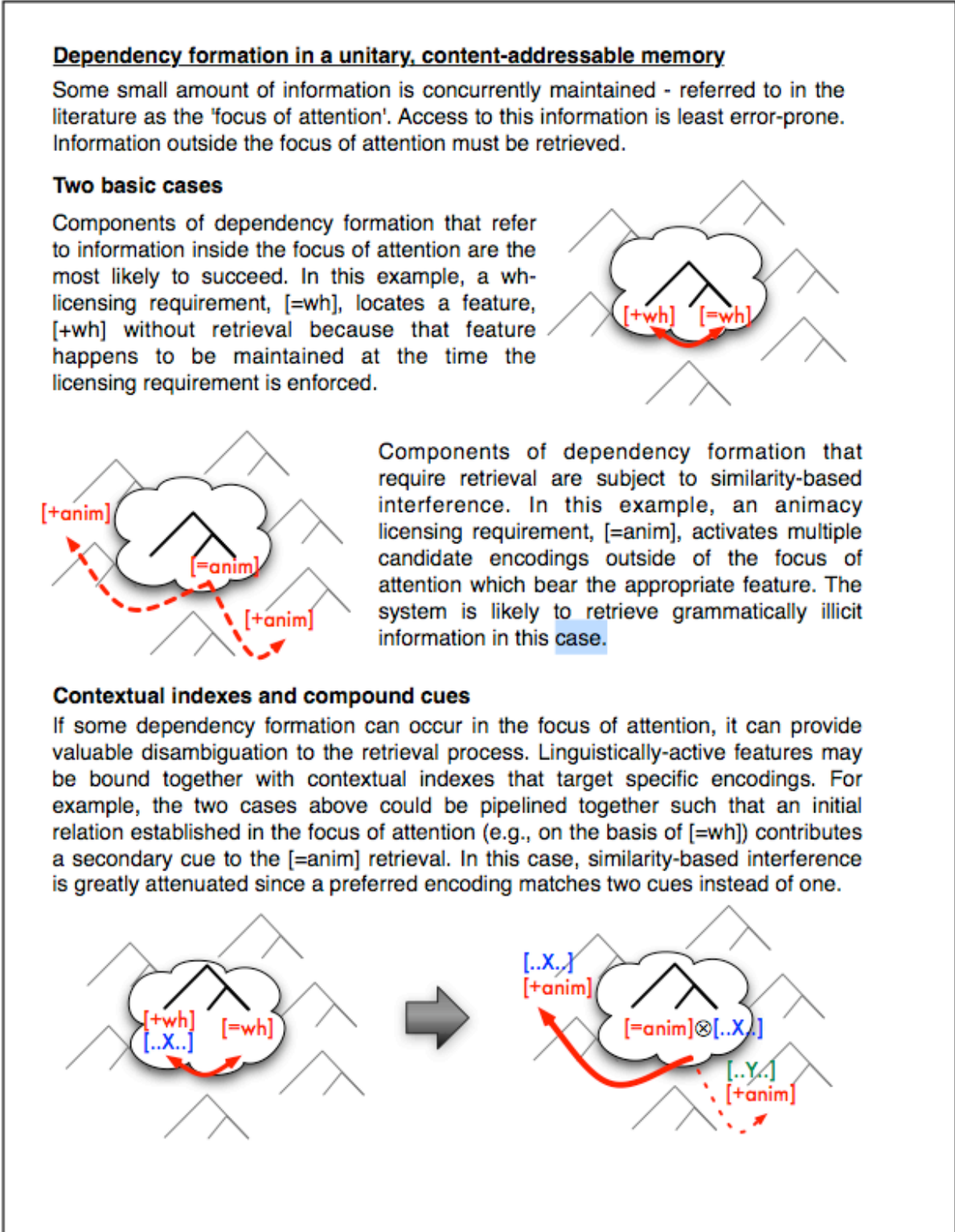


Figure 1 Dependency formation in a unitary, content-addressable memory

#### 4.2 *Implications for islandhood*

Thus far I have been mostly concerned with clarifying the memory processes involved in constructing a filler gap dependency. Now it is time to return to the original question: can any properties of the memory system contribute to an explanation of island effects? In other words, is the inability of filler-gap dependencies to cross certain boundaries attributable to an inability to construct the dependency in realtime because of memory failures? I emphasize that I am restricting the question of interest to the contribution of the working memory system to parsing failure. Under a multi-componential account, like Kluender & Kutas (1993) or Kluender (2004), many difficulties might combine together and those difficulties may relate to other properties of the system.

I think that the totality of the evidence suggests that any failures must lie in the retrieval process -- that is, any failures must stem from the parser's inability to *retrieve* filler information at some point during dependency completion. At present the evidence for maintenance in normal, uneventful long-distance dependency completion is too weak to attribute variation in acceptability among structures to that cause. Pernicious effects of retrieval interference, on the other hand, and the importance of well-chosen retrieval structures, are abundantly clear throughout both the memory literature and the psycholinguistics literature.

Retrieval interference is thus the strongest candidate to underlie any hypothetical memory failures in long-distance dependency completion. However, at the same time, it seems highly unlikely that retrieval inference is ever severe enough in sentence comprehension to guarantee actual retrieval failures. There are at least four reasons to believe this is the case. The first is simply a theoretical observation, and the next three are

empirical. The theoretical argument concerns expert performance and practiced skills, the cases Ericsson & Kintsch (1995) were worried about, and the cases for which robustness to retrieval interference is well documented. It would seem odd that sentence comprehension is not well practiced enough to allow comprehenders to develop retrieval structures to overcome any interference which island configurations might generate (for example, the strategy I outline in Figure 1). To see that this is likely not the case, we only need consider the phenomenon of resumption. Many languages allow resumptive pronouns in the place of gaps, and, interestingly, the use of resumptive pronouns often allows the grammar to form an unbounded long-distance dependency that would otherwise violate an island constraint (McCloskey, 2002). Though it has sometimes been speculated that resumptive pronouns might aid in the retrieval of filler material, much recent evidence indicates that resumption does not obviously improve comprehension performance on (non-island) long-distance dependences (Alexopoulou & Keller, 2007). If resumption does not enhance the retrievability of the filler, then the existence of island-crossing dependencies with resumptive pronouns strongly suggest that it is possible to retrieve filler material inside an island when it is a grammatical option<sup>15</sup>.

The second reason to doubt that retrieval interference leads to significant retrieval failures comes from the studies that directly document its existence. Consider, in particular, Van Dyke & McElree (2006), an experiment which explicitly introduces memory load items.

---

<sup>15</sup> Some very recent evidence by Yoshida, Kazanina, Pablos & Sturt (2011) has shown that other unbounded dependencies are readily constructed inside islands, as long as they are not A-bar dependencies. In particular, they show that the resolution of cataphoric dependencies (in which a pronoun is introduced prior to any accessible discourse referent) proceeds actively into relative clause islands. There are clearly differences in the representation of these two dependency types, but the functional demands on the comprehender are similar.

In that experiment, comprehension accuracy declined only 4% between interfering and non-interfering conditions. Recall accuracy -- how well items on the load list were recalled -- only declined 2%, a non-significant effect. Interestingly, in experiments documenting interference effects for subject-verb attachment, comprehension decrements due to interference were comparable in magnitude (Van Dyke & Lewis, 2003, Van Dyke, 2007, Wagers, 2008). In those experiments, the interference manipulation was to vary the number of subject phrases embedded inside the matrix subject and the measure was performance on the matrix verb. Yet despite the existence of comparable interference effects on subject-verb attachment, I am not aware of any proposed grammatical constraints on the relative complexity of pre-verbal subjects in English that are comparable in their severity to island constraints. It is true that other kinds of dependency formation show much larger interference effects than the studies above: in particular, higher rates of parsing or interpretation error can be found in establishing subject-verb agreement (Wagers, Lau & Phillips, 2009) or resolving reflexive anaphora (eventually) (Sturt, 2003). However I take the Van Dyke & McElree (2006) study, with its relatively low rates of interference, as the most relevant to the question at hand since it deals specifically with a *wh*-dependency.

The third reason to doubt that retrieval interference underlies islandhood comes from recent studies seeking to correlate measures of WM efficiency/capacity with acceptability ratings (Sprouse, Wagers, & Phillips, 2011, Tokimoto, 2009). Sprouse, Wagers, & Phillips (2011) attempted to correlate scores from two kinds of WM-sensitive test with the acceptability decrements island-violating sentences earned in ratings tasks: performance on an *n*-back task, as well as performance on a word span task (immediate

free recall). The results were unequivocal: there was no useful covariation between memory scores and the severity of island violations. Of course it is always possible that the wrong index of WM was selected - as a host are available - and other candidate WM indices should be tested. Yet it is important to note that the simple recall task Sprouse and colleagues employed shares considerable variance (40%) with more complex span tasks, like operation span (Kane, et al., 2004). Therefore it seems less likely that merely choosing another WM index will uncover significant covariation. Pursuing more direct measures of WM operation (e.g., the techniques in Section 2) is preferable in any case.

#### 4.3. *Closing*

In this chapter I have attempted to interpret existing psycholinguistic theory and data in a way that clarifies the time course of long-distance dependency formation and aligns it with the component processes of the working memory system. I hasten to add that our empirical database is rich in many aspects, but in others it still requires us to make a few educated guesses. Where we are data-rich concerns word-by-word time course questions: what kinds of dependencies are formed, and at what landmark in the sentence. Such information has been derived primarily from carefully designed reading studies (with both behavioral and electrophysiological measures). Where we remain theory- and data-poor is in charting the contents of information maintained not only from past events but also information about expected future events. Such information requires relatively resource intensive probe recognition experiments, implemented as cross-modal lexical priming with multiple interruption or in speed-accuracy trade studies. However I suspect that coming to grips with the contents of the relevant internal representations will require

not only that we do more of the relevant kinds of experiments. We must also push more strongly on the important distinction between the underlying events of comprehension, i.e., the instantiation and manipulation of internal linguistic representations, and the more indirect events, i.e., the incoming words, which supply the system with evidence for those representations. The active formation of long-distance dependencies reminds us that the two need not align in time.



## 5. References

- Alexopoulou, T. & Keller, F. (2007). Locality, cyclicity and resumption, at the interface between the grammar and the human sentence processor. *Language*, 83, 110-160.
- Anderson, M. C., & Neely, J.H. (1996). Interference and inhibition in memory retrieval. In E.L. Bjork & R.A. Bjork (Eds.), *Handbook of Perception and Memory. Vol. 10: Memory* (pp. 237-313). San Diego, Academic Press.
- Aoshima, S., Phillips, C., & Weinberg A. (2004). Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language*, 51, 23-54.
- Baddeley, A.D. (1986). *Working Memory*. Oxford, Clarendon Press.
- Boland, J., Tanenhaus, M., & Garnsey, S. (1995). Verb argument structure in parsing and interpretation, Evidence from wh-questions. *Journal of Memory & Language*, 34, 774-806.
- Bresnan, J. (2001). *Lexical-Functional Syntax*. Oxford, Blackwell.
- Broadbent, D. E. (1958). *Perception and communication*. New York, Oxford UP.
- Brody, M. (1995). *Lexico-Logical Form, A radically minimalist theory*. Cambridge, Mass., The MIT Press.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, Mass., The MIT Press.
- Clahsen, H. & Featherston, S. (1999). Antecedent priming at trace positions, evidence from German scrambling. *Journal of Psycholinguistic Research*, 28, 415-437.
- Clark, S. & Gronlund, S. (1996). Global matching models of recognition memory, How the models match the data. *Psychonomic Bulletin & Review*, 3(1), 37-60.
- Conklin, K., Koenig, J.P., & Mauener, G. (2004). The role of specificity in the lexical encoding of participants. *Brain and Language*, 90, 221-30.

- Cowan, N. (1995). *Attention and Memory, An Integrated Framework*. New York, Oxford UP.
- Cowan, N. (2001). The magical number 4 in short-term memory, a reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87-185.
- Cowan, N. (2005). *Attention and memory, An integrated framework*. Oxford, England, Oxford University Press.
- Conway, A.R. A., Jarrold, C., Kane, M. J., Miyake, A., & Towse, J. (2007). *Variation in Working Memory*. Oxford University Press.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks, A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12, 769-786.
- Crain, S., & Fodor, J.D. (1985). How can grammars help parsers? In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language parsing*. Cambridge, Cambridge University Press.
- De Vincenzi, M. (1991). *Syntactic parsing strategies in Italian*. Dordrecht, Kluwer Academic.
- Deane, P. (1991). Limits to attention, A cognitive theory of island phenomena. *Cognitive Linguistics*, 2(1), 1-63
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211-245.
- Fiebach, C.J., Schlesewsky, M., & Friederici, A.D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing, The processing of German WH-questions. *Journal of Memory and Language*, 47, 250-272.
- Fodor, J. (1978). Parsing Strategies and Constraints on Transformations. *Linguistic Inquiry*, 9(3), 427-73.

- Fodor, J.A., Bever, T.G. & Garrett, M.F. (1974). *The psychology of language, an introduction to psycholinguistics and generative grammar*. New York, McGraw-Hill
- Frank, R. (2002). *Phrase Structure Composition and Syntactic Dependencies*. Cambridge, Mass., The MIT Press.
- Frazier, L. (1987). Syntactic processing, Evidence from Dutch. *Natural Language and Linguistic Theory*, 5, 519–560.
- Frazier, L. & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, 4, 93-126.
- Frazier, L. & Fodor, J.D. (1978). The sausage machine, A new two-stage parsing model. *Cognition*, 6(4), 291-325.
- Frazier, L., & Flores D'Arcais, G.B. (1989). Filler-driven parsing, a study of gap filling in Dutch. *Journal of Memory of Language*, 28, 331–44.
- Garavan, H. (1998). Serial attention within working memory. *Memory and Cognition*, 26, 263-276.
- Garnsey, S.M., Tanenhaus, M.K., & Chapman, R.M. (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research*, 18, 51–60.
- Gibson, E. (1998). Syntactic complexity, Locality of syntactic dependencies. *Cognition*, 68, 1–76
- Gibson, E. (2000). The dependency locality theory, a distance-based theory of linguistic complexity. In Y. Miyashita, A. Marantz, & W. O'Neil (Eds.), *Image, Language, Brain*. Cambridge, Mass, MIT Press.

- Gibson, E. & Thomas, J. (1999). Memory Limitations and Structural Forgetting, The Perception of Complex Ungrammatical Sentences as Grammatical. *Language and Cognitive Processes*, 14(3), 225-248.
- Gibson, E. & Grodner, D.J. (2005). Consequences of the Serial Nature of Linguistic Input for Sentential Complexity. *Cognitive Science*, 29, 261-291.
- Gillund, G. & Shiffrin, R.M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, 91(1), 1-67.
- Givón, T. (1979). *On understanding grammar*. New York, Academic Press.
- Hayes, J.R.M. (1952). Memory span for several vocabularies as a function of vocabulary size. In *Quarterly Progress Report*. Cambridge, Mass., Acoustics Laboratory, Massachusetts Institute of Technology.
- Hinton, G.E. (1989). Connectionist Learning Procedure. *Artificial Intelligence*, 40, 185-234.
- Hintzman, D.L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, 95, 528-551.
- Hofmeister, P. & Sag, I.A. (2010). Cognitive constraints and island effects. *Language*, 86(2), 366-415.
- Huang, C.T.J. (1981). *Logical Relations in Chinese and the Theory of Grammar*, Doctoral dissertation, MIT, Cambridge, Mass.
- Jackendoff, R.S. & Culicover, P. (1971). A Reconsideration of Dative Movements. *Foundations of Language*, 7(3), 397-412.
- James, W. (1890). *The Principles of Psychology*, 2 vols. New York, Holt.
- Jonides, J., R.L. Lewis, D.E. Nee, C.A. Lustig, M.G. Berman & Moore, K.S. (2008). The Mind and Brain of Short-Term Memory. *Annual Review of Psychology*, 59(1), 193-224.

- Kaan, E. (1997). Processing Subject-Object Ambiguities in Dutch. Doctoral dissertation, University of Groningen.
- Kaan, E., A. Harris, E. Gibson & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, 15, 159-201.
- Kane, M.J., Hambrick, D.Z., Tuholski, S.W., Wilhelm, O., Payne, T.W., & Engle, R.W. (2004). The generality of working memory capacity, a latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology, General*, 133, 189-217.
- King, J., & Just, M.A. (1991). Individual differences in syntactic processing, the role of working memory. *Journal of Memory and Language*, 30, 580-60.
- King, J.W. & Kutas, M. (1995). Who Did What and When? Using Word- and Clause-Level ERPs to Monitor Working Memory Usage in Reading. *Journal of Cognitive Neuroscience*, 7(3), 376-395.
- Kluender, R. (2004). Are subject islands subject to a processing account? In B. Schmeiser, V. Chand, A. Kelleher & A. Rodriguez (Eds.), *WCCFL 23 Proceedings*, 101-125. Somerville, MA, Cascadilla Press.
- Kluender, R. & Kutas, M. (1993). Subjacency as a processing phenomenon. *Language and Cognitive Processes*, 8, 573-633.
- Kohonen, T. (1977). *Associative memory, a system theoretical approach*. New York, Springer-Verlag.
- Lau, E.F., Yeung, H.H., Hashimoto, R., Braun, A. & Phillips, C. (2006). Time-course and localization of syntactic and semantic anomaly responses in sentence processing, a within-subjects fMRI/MEG design. Poster presented at the 36th Society for

Neuroscience Annual Meeting, Atlanta, GA.

Lee, M.-W. (2004). Another look at the role of empty categories in sentence processing (and grammar). *Journal of Psycholinguistic Research*, 33, 51–73

Lewis, R.L. (1996). Interference in short-term memory, The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research*, 25(1), 93-115.

McElree B. (2006). Accessing recent events. *Psychology of Learning and Motivation*, 46, 155–200.

McElree, B. (1996). Accessing short-term memory with semantic and phonological information, a time-course analysis. *Memory & Cognition*, 24(2), 173-87.

McElree, B. (1998). Attended and nonattended states in working memory, accessing categorized structures. *Journal of Memory & Language*, 38, 225–52.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory. *Journal of Psycholinguistic Research*, 29(2), 111-123.

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology, Learning, Memory & Cognition*, 27, 817–35.

McElree, B. (2006). Accessing Recent Events. *The Psychology of Learning and Motivation*, 46(6), 155-200.

McElree, B. & Doshier, B.A. (1989). Serial position and set size in short-term memory, Time course of recognition. *Journal of Experimental Psychology, General*, 118, 346-73.

McElree, B., Foraker, S. & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory & Language*, 48(1), 67-91.

McKoon, G. & Ratcliff, R. (1994). Sentential context and on-line lexical decision tasks. *Journal of Experimental Psychology, Language, Memory and Cognition*, 20, 1239-43.

- Miller, G.A. & Chomsky, N. (1963). Finitary models of language users. In R. Luce, R. Bush & E. Galanter (Eds.), *Handbook of Mathematical Psychology*, Vol II. Wiley and Sons, New York.
- Miller, G.A. (1956). The magical number seven, plus or minus two, Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
- Miyake, A. & Shah, P. (1999). *Models of working memory, Mechanisms of active maintenance and executive control*. New York, Cambridge University Press.
- Murdock, B.B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89, 609-626.
- Nairne, J. (2002). The myth of the encoding-retrieval match. *Memory*, 10(5-6), 389-95.
- Nicol, J. & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, 18, 5-19.
- Nicol, J.L., Fodor, J.D., & Swinney, D. (1994). Using cross-modal lexical decision tasks to investigate sentence processing. *Journal of Experimental Psychology, Learning, Memory & Cognition*, 20, 1229-38.
- Oberauer, K. (2002). Access to information in working memory, exploring the focus of attention. *Journal of Experimental Psychology, Learning, Memory & Cognition*, 28, 411-21.
- Oberauer, K. (2006). Is the focus of attention in working memory expanded through practice? *Journal of Experimental Psychology, Learning, Memory & Cognition*, 32, 197-214.
- Omaki, A., Lau, E., Davidson White, I., Phillips, C. (2011). Hyper-active gap filling, pre-verbal object gap creation in English filler-gap dependency processing. Manuscript in

submission.

- Öztekin, I., McElree, B., Staresina, B.P., Davachi, L. (2008). Working memory retrieval, Contributions of left prefrontal cortex, left posterior parietal cortex and hippocampus. *Journal of Experimental Psychology, General*, 138, 64-87.
- Phillips, C. (2006). The real-time status of island phenomena. *Language*, 82, 795–823.
- Phillips, C. & Wagers, M. (2007). Relating Structure and Time in Linguistics and Psycholinguistics. In G. Gaskell (Ed.), *Oxford Handbook of Psycholinguistics*. Oxford, Oxford University Press.
- Phillips, C., Kazanina, N., & Abada, S.H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, 22, 407–428.
- Phillips, C., Wagers, M. & Lau, E. (In press). Grammatical illusions and selective fallibility in real-time language comprehension. In J. Runner (Ed.), *Experiments at the interfaces, Syntax & Semantics, vol. 37*, Bingley, UK: Emerald Publications.
- Pickering, M.J. & Traxler, M.J. (2001). Strategies for processing unbounded dependencies, Lexical information and verb-argument assignment. *Journal of Experimental Psychology, Learning, Memory, and Cognition*, 27, 1401-10.
- Pickering, M.J. & Traxler, M.J. (2003). Evidence against the use of subcategorisation frequency in the processing of unbounded dependencies. *Language and Cognitive Processes*, 18, 469-503.
- Pickering, M.J., Barton, S., & Shillcock, R. (1994). Unbounded dependencies, island constraints and processing complexity. In C. Clifton, L. Frazier & K. Rayner (Eds.), *Perspectives on Sentence Processing*. London, Lawrence Erlbaum.



- Pollack, I. (1953). The assimilation of sequentially encoded information. *American Journal of Psychology*, 66, 421-435.
- Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, 116 (1), 129-156.
- Pritchett, B.L. (1991). Head position and parsing ambiguity. *Journal of Psycholinguistic Research*, 20(3), 251-270.
- Pritchett, B.L. (1992). *Grammatical competence and parsing performance*. Chicago, IL, The University of Chicago Press.
- Radó, J. (1999). Some effects of discourse salience on gap-filling. Poster presented at *The 12th Annual CUNY Conference on Human Sentence Processing*.
- Ruchkin, D.S., Johnson, R., Canoune, H., & Ritter, W. (1990). Short-term memory storage and retention, An event-related brain potential study. *Electroencephalography and Clinical Neurophysiology*, 76, 419-39.
- Rumsfeld, D.H. (Presenter). (2002, February 12). *Department of Defense News Briefing* [TV Transcript]. Washington, D.C., Department of Defense. Retrieved from <http://www.defense.gov/transcripts/transcript.aspx?transcriptid=2636>.
- Sag, I.A., Wasow, T., & Bender, E.M. (2003). *Syntactic Theory, A Formal Introduction*. Number 152 in CSLI Lecture Notes. CSLI Publications.
- Schlesewsky, M., Fanselow, G., Kliegl, R., Krems, J. (2000). The subject preference in the processing of locally ambiguous WH-questions in German. In B. Hemforth & L. Konieczny (Eds.), *German Sentence Processing*. Kluwer.
- Sekerina, I.A. (2003). Scrambling and processing, dependencies, complexities and

- constraints. In S. Karimi (Ed.), *Word Order and Scrambling*. Malden, MA, Blackwell Publishers.
- Sprouse, J., Wagers, M., Phillips, C. (In press). A test of the relation between working memory capacity and syntactic island effects. *Language*.
- Stabler, E.P. (1994). The finite connectivity of linguistic structures. In, Clifton, C., Jr., Frazier, L., Rayner, K. (Eds.), *Perspectives on Sentence Processing*. Hillsdale, NJ, Erlbaum, 303-336.
- Staub, A. (2007). The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology, Learning, Memory, and Cognition*, 33, 550-569.
- Steedman, M. (2000). *The Syntactic Process*. Cambridge, Mass., The MIT Press.
- Stowe, L. (1986). Parsing wh-constructions, evidence for on-line gap location. *Language and Cognitive Processes*, 1, 227-46
- Stowe, L.A., Tanenhaus, M.K., & Carlson, G.M. (1991). Filling gaps on-line, Use of lexical and semantic information in sentence processing. *Language and Speech*, 34, 319-34.
- Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48, 542-562.
- Sussman, R.S. & Sedivy, J.C. (2003). The time-course of processing syntactic dependencies, evidence from eye-movements. *Language and Cognitive Processes*, 18, 143-163.
- Swinney, D., Ford, M., Frauenfelder, U., & Bresnan, J. (1988). On the temporal course of gap-filling and antecedent assignment (luring sentence comprehension). In B. Grosz, R. Kaplan, M. Macken & I. Sag (Eds.), *Language and Structure Processing*. Stanford, CA, CSLI.
- Tanenhaus, M., Carlson, G., Seidenberg, M. (1985). Do listeners compute linguistic

- representations? In A. Zwicky, L. Karttunen & D. Dowry (Eds.), *Natural language parsing, Psycholinguistic, theoretical, and computational perspectives*. London and New York, Cambridge University Press.
- Tokimoto, S. (2009). Island phenomenon in Japanese and working memory, syntactic constraints independent from working memory constraints. Poster presented at the 22nd Annual CUNY Sentence Processing Conference, Davis, CA.
- Traxler, M.J. & M.J. Pickering (1996). Plausibility and the Processing of Unbounded Dependencies, An Eye-Tracking Study. *Journal of Memory and Language*, 35, 454-475.
- Tulving, E. & D.M. Thomson (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80, 352-373.
- Van Dyke, J.A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology, Learning, Memory, and Cognition*, 33(2), 407-30.
- Vasishth, S. & Lewis, R.L. (2005). An Activation-Based Model of Sentence Processing as Skilled Memory Retrieval. *Cognitive Sciences*, 29, 375-419.
- Vasishth, S., Lewis, R., Kern, S., & Suckow, K. (2010). Short-term forgetting in sentence comprehension, Crosslinguistic evidence from verb-final structures. *Language and Cognitive Processes*, 25(4), 533-567.
- Wagers, M. (2008). *The structure of memory meets memory for structure in linguistic cognition*. Doctoral dissertation, UMD, College Park, MD.
- Wagers, M., Lau, E., & C. Phillips (2009). Agreement attraction in comprehension, representations and processes. *Journal of Memory and Language*, 61, 206-237.

- Wagers, M. & McElree, B. (2011). Memory for linguistic features, evidence from the dynamics of agreement. Unpublished manuscript.
- Wagers, M. & Phillips, C. (2009). Multiple dependencies and the role of the grammar in real-time comprehension. *Journal of Linguistics*, 45, 395-433.
- Wagers, M. & Phillips, C. (2011). Going the distance, memory and decision-making in active dependency construction. Manuscript submitted.
- Wanner, E., & Maratsos, M. (1978). An ATN Approach to Comprehension. In M. Halle, J. Bresnan, & G.A. Miller, (Eds.), *Linguistic Theory and Psychological Reality*, Cambridge, MA, MIT Press.
- Weinberg, A. (1992). Parameters in the Theory of Sentence Processing, Minimal Commitment Theory goes East. *Journal of Psycholinguistic Research*, 22(3), 339-64.
- Wickelgren, W.A. (1976). Network strength theory of storage and retrieval dynamics. *Psychological Review*, 83, 466-78.
- Yngve, V. (1961). The depth hypothesis. In R. Jakobson (Ed.), *Structure of language and its mathematic aspects*. Providence, RI, American Mathematical Society.
- Yoshida, M., Kazanina, N., Pablos, L., & Sturt, P. (2011). On the origin of islands. Submitted manuscript.