Memory Limitations and Structural Forgetting: The Perception of Complex Ungrammatical Sentences as Grammatical

Edward Gibson

Massachusetts Institute of Technology, Cambridge, USA

James Thomas

Carnegie Mellon University, Pittsburgh, USA

Results from an English acceptability-rating experiment are presented which demonstrate that people find doubly nested relative clause structures just as acceptable when only two verb phrases are included instead of the grammatically required three. Furthermore, the experiment shows that such sentences are acceptable only when the intermediate verb phrase is omitted. A number of specific accounts of forgetting are considered. Two early proposed theories of this effect, the disappearing syntactic nodes hypothesis (Frazier, 1985) and the least recent nodes hypothesis (Gibson, 1991), are not consistent with the experimental results. The results, together with other acceptability patterns, suggest that the representations that are retained (and subsequently forgotten) in processing sentences consist of the lexical word-strings processed thus far. Three possible accounts of the results are considered: (1) the high memory cost pruning hypothesis within the framework of Gibson (1998); (2) a recency/primacy account; and (3) a connectionist account (Christiansen & Chater, in press).

INTRODUCTION

It has long been observed that *nested* sentence structures like (1) are very difficult to understand (Bever, 1970; Chomsky & Miller, 1963; Cowper,

---

1976; Gibson, 1991, 1998; Kimball, 1973; Lewis, 1993, 1996; Miller & Chomsky, 1963; Miller & Isard, 1964; Stabler, 1994):[1]

(1) # [$_s$ The patient [who [$_s$ the nurse [who [$_s$ the clinic had hired]] admitted]] met Jack].

A syntactic category, *A*, is said to be nested (or *centre-embedded*) within another category *B* if *B* contains *A*, a constituent to the left of *A*, and a constituent to the right of *A*. In (1), the relative clause (RC) "who the nurse … admitted" is nested within the sentence "the patient … met Jack". Furthermore, a second RC "who the clinic had hired" is nested within the first embedded sentence "the nurse … admitted". In contrast, structures that are left- or right-branching are much easier to understand than nested structures. For example, the right-branching structure in (2) has the same meaning as its nested counterpart in (1) at the level of thematic structure, but it is much easier to understand.

(2) Jack met the patient [who was admitted by the nurse [who the clinic had hired]].

According to most theories, the difficulty associated with processing nested structures is due to resource limitations. For example, according to Kimball (1973), the limitation on resources is quantified in terms of the number of partially processed clauses that are kept in memory at any point in processing a sentence, such that the processor can keep track of at most two incomplete clauses at any one time. Processing the first subject "the patient" in (1) causes the initiation of one clause which will be completed when the verb "met" and its immediate dependents are located in the input string. Processing the following two subjects "the nurse" and "the clinic" of the embedded RCs causes the initiation of two further clauses, resulting in a total of three partially processed clauses, which is more than the resource capacity. In contrast, there is never more than one incomplete clause in processing the right-branching structure in (2), so this sentence is processed without difficulty.

Under the assumption that some linguistic structures require more resources than are available, an interesting question that arises is what the processor does at points near or beyond the limits of resource capacity. Does the processor simply stop processing as soon as its resource capacity is exceeded? Alternatively, does the processor selectively forget some of the partially processed information in order to continue? An interesting source of evidence relevant to this question comes from the processing of

---

[1] Sentences that cause extreme processing difficulty are prefixed with the symbol "#".

complex, ungrammatical items like (3), which lack the second verb phrase (VP) in a doubly nested RC structure.[2] [An example like (3) is discussed in Frazier (1985), where the acceptability observation is attributed to Janet Fodor.]

(3) $\sqrt{}$* The patient who the nurse who the clinic had hired met Jack.

Surprisingly, intuitions suggest that examples like (3) are perceived as grammatical and perhaps somewhat easier to process than examples like (1), despite the fact that they are missing a VP [''admitted'' in (1)] that is required by the syntactic rules of English. Furthermore, Fodor's intuition is that it is the second VP that must be left off in order to give the perception of a grammatical sentence. We will refer to the perceived grammaticality of sentences like (3) as the ''missing VP'' effect.

The relative acceptability of (3) provides a partial answer to the question posed earlier. It suggests that the processor continues processing at points of high complexity [e.g. during the processing of ''the clinic had hired'' in (1) and (3)] after possibly forgetting some structures that it built earlier. One consequence of this ''structural forgetting'' is that leaving off one of the required VPs from a doubly nested RC structure gives the impression of a grammatical sentence.

This paper investigates the nature of structural forgetting using a complexity rating experiment. Two structural forgetting theories that have been proposed in the literature are the disappearing syntactic nodes hypothesis (Frazier, 1985) and the least recent nodes hypothesis (Gibson, 1991). A third possibility follows from the resource theory proposed in Gibson (1998). Each of these theories is summarised later, and their predictions are then evaluated with respect to the complexity-rating experiment. Further possibilities are evaluated within the General Discussion. In addition to providing information about how the processor operates at points of high complexity, understanding the nature of structural forgetting also has the potential to provide information about the kinds of representations that the processor operates on. This point is elaborated in the General Discussion with respect to the experimental results.

The disappearing syntactic nodes hypothesis

The earliest proposal of a process of structural forgetting is Frazier's (1985) *disappearing syntactic nodes hypothesis*, which was intended to account for

---

[2] We will prefix sentences that are ungrammatical but yet seem grammatical with the symbols ''$\sqrt{}$*''.

the intuition that it is the middle VP that is omitted in processing a sentence like (4), which is the same as (3) but with the lexical RC pronouns ("who") omitted.

(4) $\sqrt{*}$ The patient the nurse the clinic had hired met Jack.

(5) The disappearing syntactic nodes hypothesis:

> At points of high complexity, syntactic nodes will tend to be forgotten if they dominate no lexical material or only dominate material that has already been semantically combined (Frazier, 1985, p. 178).

According to Frazier, a noun phrase (NP) is semantically combined when it receives a thematic role or is coindexed with an RC pronoun that receives a thematic role. Thus, the NP "the clinic" in (4) is semantically combined with the verb "hired", the verb that assigns it a thematic role. The NP "the nurse" is also semantically combined with the verb "hired", because "hired" assigns a thematic role to a nonlexical RC pronoun coindexed with "the nurse".

The operation of the disappearing syntactic nodes hypothesis as applied to (4) proceeds as follows. First, Frazier hypothesises that a large quantity of resources are necessary at the point of processing the most embedded subject NP "the clinic" in the doubly embedded RC structure. This is true under Frazier's (1985) complexity theory as well as all other theories of nesting complexity. See Fig. 1 for a diagram of the tree structure built at this point, with each predicted VP represented by a question mark in the figure. The disappearing syntactic nodes hypothesis does not apply at this point in the parse, because there are no structures that have been semantically combined and all structures dominate some lexical material. Processing continues, and the first VP "had hired" is integrated into the structure. The verb "hired" assigns one thematic role to "the clinic" and a second thematic role to a nonlexical RC pronoun coindexed with "the nurse", thereby completing the RC modifying "the nurse". Because the NPs "the nurse" and "the clinic" are semantically combined at this point, the structures for the NP "the nurse" and its modifying RC are pruned away from the structure according to the disappearing syntactic nodes hypothesis. The predicted RC structure following "the patient" can also be pruned, because it no longer dominates any lexical material. The matrix sentence node remains, dominating the matrix subject NP "the patient". This structure cannot be pruned, because it includes structures representing lexical items that have not yet been semantically combined. The verb "met" is then interpreted as the matrix verb, and the sentence is perceived as grammatical.

Although this theory captures the desired missing VP effect in (4), a substantial number of stipulations are required. In particular, the definition
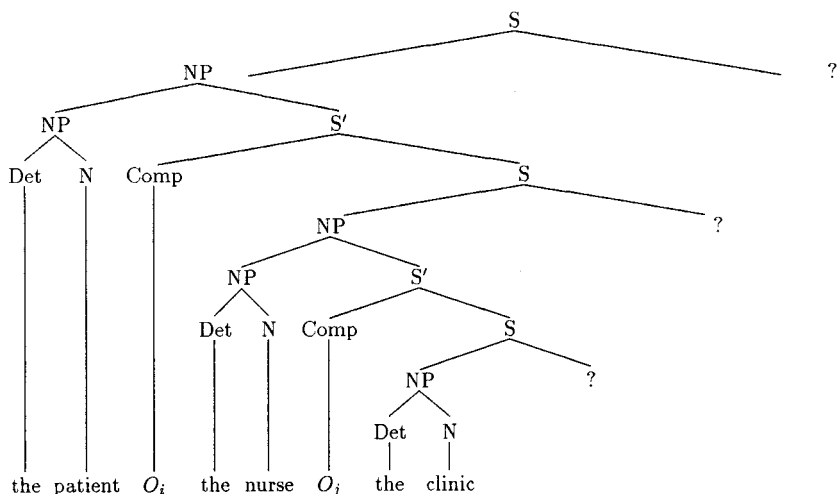
FIG. 1.   The syntactic structure for the partial sentence "The patient the nurse the clinic …". Nonlexical wh-pronouns ("operators") are represented as "O_i" and "O_j".

of semantic combination is necessarily disjunctive, stating that an NP is semantically combined if it either receives a thematic role directly or is coindexed with an RC pronoun that receives a thematic role. If the direct thematic role assignment component of the definition were omitted, then the NP "the clinic" would not be forgotten. If the indirect RC pronoun component of the definition were omitted, then the NP "the nurse" would not be forgotten. And if either of these components were not forgotten, then all of the VPs would be required to form a grammatical sentence, and the missing VP effect would not result.

In addition to the fact that the theory is stipulative, the theory has a number of empirical problems. First, Frazier's account of the acceptability of (4) relies crucially on the fact that the pronoun in the first RC is nonlexical. If this pronoun is lexical [e.g. "who" as in (3)] then the disappearing syntactic nodes hypothesis would not apply to delete the clause after "patient", because this clause would contain a lexical element: The pronoun "who". Thus, the disappearing syntactic nodes hypothesis predicts that there should be no missing VP effect in examples like (3), in which the wh-pronouns are lexically realised. Contrary to prediction, intuitions suggest that the missing VP effect does apply in examples like (3). The experiment reported below tests this prediction.

A second related problem with the disappearing syntactic nodes hypothesis is that it predicts that examples like (6) should be as acceptable as examples like (4):

(6) * The patient who the nurse who the clinic had hired was smiling met
     Jack.

After processing "hired", the structure for the embedded NP "the nurse
who the clinic had hired" can be forgotten according to the disappearing
syntactic nodes hypothesis, leaving the structure for "the patient who". It
should then be possible to attach an intransitive VP such as "was smiling"
as the predicate for the outer RC, as in the sentence "the patient who was
smiling met Jack". Intuitions suggest that this prediction is incorrect: (6) is
much less acceptable than (1), (3) or (4).

The least recent nodes hypothesis

Gibson (1991) provided an alternative formulation of structural forgetting:

(7) The least recent nodes hypothesis:

     If a structure requires more processing load than the available
     capacity, then selectively remove (forget) nodes directly dependent
     on the least recent words in the input string until the load associated
     with the structure is lowered below the desired threshold (Gibson,
     1991, p. 168).

The effect of this hypothesis is to prune away structures associated with
the syntactic nodes created furthest back in the parse, in the face of high
syntactic complexity. This hypothesis therefore makes a different predic-
tion from Fodor's intuition about the missing VP effect: It predicts that the
final verb phrase is the one that is preferentially omitted, not the middle
one. Note that the fact that (3) and (4) are acceptable is potentially
consistent with forgetting either of the first two NPs, because the final verb
phrase "met Jack" is a plausible predicate for either the first NP "the
patient" or the second NP "the nurse". Thus, the acceptability of these
examples does not distinguish the hypothesis that the first NP is forgotten
(the least recent nodes hypothesis) from the hypothesis that the second NP
is forgotten (the disappearing syntactic nodes hypothesis). The experiment
presented here addresses these alternative predictions.

The high memory cost pruning hypothesis

Although most theories of syntactic complexity associate the same
quantity of memory cost with each incomplete syntactic relationship
relevant in the theory, Gibson's (1998) theory—the syntactic prediction
locality theory (SPLT)—associates different costs with different incom-
plete syntactic relationships, depending largely on the distance since the
incomplete syntactic relationship was first identified (cf. Hawkins, 1994;
Joshi, 1990; Just & Carpenter, 1992; Rambow & Joshi, 1994). This property

of the SPLT leads naturally to a theory of structural forgetting according to which the most costly predictions are forgotten first:

(8) The high memory cost pruning hypothesis: At points of high memory complexity, forget the syntactic prediction(s) associated with the most memory load.

There are two components to the SPLT: A memory cost component and an integration cost component. According to the SPLT memory cost component, each syntactic head that is required to complete the current input string as a grammatical sentence is associated with a memory cost. The processor is therefore making predictions for required syntactic heads, each of which is associated with a cost. The total memory cost at a word in a partially processed input sentence is then calculated by summing together the costs for all the required syntactic heads at that point. This theory requires a hypothesis about the set of syntactic heads comprising sentences. For simplicity, a syntactic theory with a minimal number of functional categories is assumed, such as Head-driven Phrase Structure Grammar (Pollard & Sag, 1994) or Lexical Functional Grammar (Bresnan, 1982). Under these theories, the minimal number of syntactic head categories in a sentence is two: A noun for the subject, and a verb for the predicate. If words are encountered that necessitate other syntactic heads to form a grammatical sentence, then these heads are also predicted, and additional memory load is incurred. The statement of the SPLT's memory cost for each required head is given in (9):

(9) Syntactic prediction memory cost:
    a. The prediction of the top-level matrix argument predicate, V_0, is associated with no memory cost.
    b. For each required syntactic head C_i other than V_0, associate a memory cost of $M(n)$ memory units (MUs), where $M(n)$ is a monotone increasing function and n is the number of new discourse referents that have been processed as C_i was initially predicted.

The top-level matrix argument predicate can be initially thought of as the structurally highest (matrix) predicate in a sentence, where the predicate in a sentence is usually a verb in English. The definition of matrix argument will be discussed in more detail later. For the purposes of the high memory cost pruning hypothesis, the important claims of the SPLT are: (a) that the prediction of the top-level predicate is cost-free; and (b) that all other syntactic predictions require an increasing quantity of memory resources as determined by the distance since the predictions were first made. Motivation for the claim that syntactic memory cost increases with distance comes from the short-term memory literature, where it has been demonstrated that it is harder to retain items (e.g.

unrelated words or digits) in short-term memory as more interfering items are processed (see e.g. Waugh & Norman, 1965; see Lewis, 1996, for a recent summary of relevant results). According to the memory cost hypothesis in (9), intervening elements which cause substantial integration and memory cost increments are words introducing new discourse referents: NPs (object referents) and the main verbs of VPs (event referents). This is a simplification of the distance-based cost hypothesis in Warren and Gibson (1998), according to which the cost is incremented depending on the difficulty of accessing a referent in the discourse (Garrod & Sanford, 1977, 1982, 1994; Haviland & Clark, 1974; Haliday & Hasan, 1976; Warren & Gibson, 1998). Focused entities or individuals, which are usually referred to with pronouns, are highly accessible, so they require a small quantity of resources to access. Nonfocused entities or individuals in the discourse require more resources to access. Such NPs are usually referred to using proper names and definite descriptions. Elements new to the discourse, which are usually introduced using indefinite NPs, require the most resources because they must be constructed in the discourse model. We will follow Gibson (1998) in assuming a simplified version of the distance cost such that only the processing of new discourse referents causes a memory cost increment. Thus, processing an NP which is new to the discourse causes a memory cost increment, but processing a pronoun—which refers to an entity already established in the discourse—does not.

The SPLT memory cost theory accounts for the difficulty in processing nested structures in a variety of constructions crosslinguistically. For example, the large complexity difference between singly nested RC structures, as in (10a), and doubly nested RC structures, as in (10b), is accounted for straightforwardly:

(10)a. The professor who$_i$ the scientist collaborated with e_i had advised the student who$_j$ e_j copied the article.

   b. The student who$_j$ the professor who$_i$ the scientist collaborated with e_i had advised e_j copied the article.

It is hypothesised that a large factor contributing to the intuitive judgement of a sentence's complexity is the maximal memory cost incurred at any point during its processing. The point of maximal complexity in processing the singly nested structure in (10a) occurs at the point of processing the noun "scientist". At this point, there are three heads that are required to complete the partial sentence grammatically: (i) the top-level verb; (ii) a verb to head the RC; and (iii) an NP empty-category to be coindexed with the RC pronoun "who". The matrix verb is assumed to be cost-free according to (9). The other two predictions were made at the point of processing the RC pronoun "who", and one new discourse referent, the NP "the scientist", has been processed since each of these predictions was

made, resulting in a cost of $M(1)$ MUs for each prediction. The point of maximal complexity in processing the doubly nested structure in (10b) also occurs at the point of processing the noun "scientist", but the cost is much greater here. In particular, five heads are required to complete this partial sentence grammatically: (i) the top-level verb (zero cost); (ii) a verb to head the first RC [$M(2)$ MUs, corresponding to two new discourse referents processed since the prediction was first made: "the professor" and "the scientist"]; (iii) an NP empty-category to be coindexed with the first RC pronoun [also $M(2)$ MUs]; (iv) a verb to head the second RC [$M(1)$ MUs, corresponding to one new discourse referent processed since the prediction was first made: "the scientist"]; and (v) an NP empty-category to be coindexed with the second RC pronoun [also $M(1)$ MUs]. Thus, the total SPLT memory cost at this point is $2M(2) + 2M(1)$ MUs, which is much greater than the maximal memory cost in processing the singly embedded structure, and the contrast is accounted for.

Empirical evidence for the discourse-based distance metric is provided by Warren and Gibson (1998) who used a complexity rating questionnaire to show that doubly nested RC structures are easier to process when a first- or second-person pronoun is in the subject position of the most embedded RC, as in (11), as compared with a similar structure in which an NP introducing a new object into the discourse is in the subject position of the most embedded clause, as in (10b) (cf. Bever, 1970; Kac, 1981):

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

This observation can be accounted for in the SPLT framework if the memory increment for a predicted category is larger for new discourse referents than for referents that are already part of the current discourse, such as first- or second-person pronouns. (It is assumed that the current discourse always includes a speaker/writer and a hearer/reader.) The maximal complexity in (10b) is $2M(2) + 2M(1)$ MUs, at the most embedded subject "the scientist". If there is no memory cost increment for referents that are already part of the current discourse (such as "I" and "you"), then the memory cost at the point of processing "I" in (11) is only $2M(1) + 2M(0)$ MUs, corresponding to one fewer new discourse referent processed for each predicted head than at the corresponding point in (10b). The memory cost at all other parse states in (11) is also lower than in (10b). As a result, the maximal memory cost required to process a doubly nested RC structure with a new referent in its most embedded subject position is greater than that required to process a similar structure with an old referent in its most embedded subject position.

The conceptual motivation for the claim that there is no memory cost associated with predicting the top-level matrix argument predicate is that

most utterances are headed by a predicate, with the consequence that the processor usually expects a predicate before a sentence has been initiated. This prediction may therefore be built into the processor. As a result, the processor predicts a predicate as a sentence is being initiated without any knowledge of the initial words in a sentence. Other syntactic heads, such as those initiating an RC, cannot be predicted ahead of time in this way, so the processor makes these predictions on-line, as lexical items are processed.

Empirical support for the claim that there is no memory cost associated with predicting the top-level matrix argument predicate is provided by Gibson and Thomas (1997), who used questionnaire complexity ratings to compare a pre-matrix-verb doubly nested RC construction to the same construction following the matrix verb. To keep the meaning the same in the two constructions, the voice was changed from active in one to passive in the other, balanced across the items:

(12)a. Pre-verbal: The administrator who the intern who the nurse supervised had bothered lost the medical reports.

  b. Post-verbal: The medical reports were lost by the administrator who the intern who the nurse supervised had bothered.

No complexity difference was observed between the two conditions. The lack of a complexity effect was likely not due to task insensitivity, because many other comparisons in the same experiment yielded highly significant differences.

The maximal memory cost in each of the constructions in (12) occurs at the point of processing the most embedded RC subject in each: "the nurse". The only difference in category requirements at this point is that (12a) still requires the matrix verb for the sentence, whereas this requirement has already been satisfied in (12b). Thus, the lack of a difference between pre- and post-verbal RC structures can only be compatible with a predicted-category-based framework like the SPLT if the prediction of the matrix verb is of low cost relative to the costs of the other predictions. Hence, Gibson assumes that the prediction of the matrix verb is not increasing in cost like other predictions, but is cost-free. (The available empirical data is also consistent with the hypothesis that the prediction of the matrix verb has some cost, as long as it is less than the costs for other predictions, so that its effects would not be so easily measurable.) Hakes, Evans, and Brannon (1976) report similar results with respect to pre- and post-verbal singly embedded RCs.

The assumption that the prediction of the matrix argument predicate is cost-free requires an additional assumption about clause-based processing. In particular, it cannot be only the matrix predicate whose prediction is

cost-free, because this would incorrectly predict that a sentence like (13b) should be more complex than a sentence like (13a):

(13) a. The reporter that the senator attacked ignored the president.
    b. The editor said that the reporter that the senator attacked ignored the president.

Sentence (13a) contains an RC modifying the subject NP of a sentence, and (13b) consists of (13a) in an embedded context. If only the prediction of the matrix verb were cost-free, then the prediction of "said" in (13b) would be cost-free, but the prediction of the embedded verb "ignored" would not. The maximal memory cost would therefore be greater in (13b) than in (13a), because the prediction of "ignored" is cost-free in (13a). Contrary to prediction, there is no complexity difference between the two. As a solution to this potential problem, Gibson proposed that the processor is clause-based, shunting completed clauses out of working memory (cf. Bever & Townsend, 1979; Caplan, 1972; Frazier & Fodor, 1978; Jarvella, 1971, 1979; Marslen-Wilson, Tyler & Seidenberg, 1978; Milsark, 1983; Roberts, Hooker, & Gibson, 1997):

(14) Clause-based closure: The initiation of a new clause causes closure of an earlier clause whose preferred dependents have been satisfied.

Under this hypothesis, a dependent is satisfied when its head is located. The matrix clause of (13b) can then be closed when the complementiser "that" is encountered following "said", because this is the head of the embedded clause. The clause-based closure idea is closest to Church's (1980) principle of A-over-A closure, which closes off an old node when a new node of the same category is identified. However, unlike A-over-A closure, clause-based closure applies only to clauses. See Church (1980) and Gibson (1991) for discussion of these and other closure principles such as Early Closure (Kimball, 1973) and Late Closure (Frazier, 1978).

In addition, it was proposed that the predicted predicate that is cost-free is the top-level matrix argument predicate remaining in the memory buffer, where a matrix argument predicate is recursively defined as either the matrix predicate or an argument of the matrix argument predicate. Consider the clause-based closure assumption with respect to (13b). Initially, the prediction of the matrix verb "said" is cost-free. When the next clause is initiated, following "that", the matrix clause is closed, and the top-level matrix argument clause becomes the clause initiated by "that". The prediction of the predicate in this clause "ignored" is therefore cost-free, and the lack of a complexity difference between (13b) and (13a) is accounted for.

Let us now explore how this theory in conjunction with the high memory cost pruning hypothesis might account for the missing VP effect. The maximal complexity of (1) and (3) (repeated here) occurs at the point of processing the most embedded NP "the clinic":

(1) # The patient who$_i$ the nurse who$_j$ the clinic had hired e_j admitted e_i met Jack.

(3) $\checkmark$* The patient who the nurse who$_j$ the clinic had hired e_j met Jack.

At this point, there are five predicted syntactic heads resulting in a total memory cost of 2M(2) + 2M(1) MUs, the same as the maximal memory cost in processing (10b). This cost is made up of: (a) no cost for the prediction of the matrix verb; (b) 2M(2) MUs for the prediction of the outer RC verb and NP-empty-category; and (c) 2M(1) MUs for the prediction of the inner RC verb and NP-empty-category. Hence, the most expensive category predictions after processing the most embedded NP "the clinic" are the two predictions associated with the first RC (the second VP to follow). Assuming that the total memory cost at this point exceeds the capacity of the system, the high memory cost pruning hypothesis predicts that these predicted categories are the ones that will be forgotten, leading to the missing VP effect in (3). Thus, the SPLT in conjunction with the high memory cost pruning assumption can account for the missing VP effect.

EXPERIMENT

This experiment was designed to test whether the missing verb phrase (VP) effect is real, and if so, which structures/dependencies are being forgotten in the processing of these complex constructions. As discussed earlier, the relative acceptability of (3) and (4) does not distinguish which of the three noun phrases (NPs) is being forgotten. It could be that the first NP is being forgotten, so that the NP "the nurse" is being taken as the subject of the verb "met". It could also be that the second NP is being forgotten, so that the NP "the patient" is being taken as the subject of the verb "met". Finally, although less plausible than the other two, it could be that the third NP is being forgotten, so that the sentence is interpreted with the nurse hiring the patient, and the patient meeting Jack. In order to determine which of these three hypotheses is correct, the experiment examines the processing of doubly nested relative clause (RC) structures whose verbs have strong semantic/pragmatic selectional restrictions for their subjects, as in (15):

(15)a. All three VPs: The ancient manuscript that the graduate student who the new card catalog had confused a great deal was studying in the library was missing a page.

    b. Missing VP_1: The ancient manuscript that the graduate student who the new card catalog was studying in the library was missing a page.

    c. Missing VP_2: The ancient manuscript that the graduate student who the new card catalog had confused a great deal was missing a page.

    d. Missing VP_3: The ancient manuscript that the graduate student who the new card catalog had confused a great deal was studying in the library.

If dependencies associated with the first (outermost) NP are being forgotten, as predicted by the least recent nodes hypothesis, then the omission of the third (outermost) VP should be the most acceptable of the missing VP conditions. Omission of either of the other VPs should result in high unacceptability, because one of the other two NPs (NP_2 or NP_3) will be forced to be interpreted as the subject of a verb which selectionally restricts against it. If dependencies associated with the second NP are being forgotten, as predicted by the high memory cost pruning assumption in conjunction with the SPLT, then the omission of the second VP should be most acceptable of the missing VP conditions. If dependencies associated with the third (innermost) NP are being forgotten, then the omission of the first (innermost) VP should be most acceptable. Finally, if the missing VP effect depends on the lack of wh-pronouns or complementisers, as Frazier's disappearing syntactic nodes hypothesis predicts, then all of the conditions lacking a VP should be worse than the one containing all three, with no difference among the conditions lacking a VP.

Method

Participants. Forty native English speakers from the Boston, Massachusetts academic community (primarily undergraduates from MIT) participated for $4.00 each.

Materials. Twelve doubly nested RC items were constructed, each with four conditions: The grammatical condition containing three VPs, and three other conditions in which one of each of the three VPs was deleted. The items were constructed such that each verb in each item had strong semantic/pragmatic selectional restrictions, so that only one of the preceding NPs was plausible as its subject. Furthermore, the second VP in each of these items was always an optionally intransitive verb. Consequently, if the second verb was taken to be the matrix verb in the missing VP_3 condition, a poor rating on this condition would be due to

selectional restriction violations between the matrix subject and this verb, and would not be due to an argument structure violation in the form of a transitive verb missing an object.

The experimental items were combined with 55 fillers (including 7 practise items) of approximately the same length as the experimental items, to form four lists. Many of the filler sentences were complex, consisting of nesting of RCs as well as other kinds of clauses. All of the filler items were grammatical sentences. The experimental items were counterbalanced across the lists so that each list contained exactly one version of every item. A list of materials is given in the Appendix.

Procedure.   The stimuli were presented in the form of a questionnaire in which the participants were asked to rate sentences for comprehensibility on a 5-point scale, where 1 was ''easy to understand'', and 5 was ''hard to understand''. Participants were instructed to read each sentence only once, and to make their rating on the basis of their first impression. Seven example sentences were presented to the participants, with discussion of possible scores for each. The first three of these examples were relatively comprehensible sentences, and the last four were much less comprehensible. One of these was an ungrammatical sentence in which a VP from a clause was omitted: ''The form was stamped by the bureaucrat who worked at the ministry where everyone who had walked strangely''. It was suggested that participants should score an example like this with a 4 or 5, because of its ungrammaticality, although no specific reason for the ungrammaticality was given. Participants completed the task within about 10–15 minutes.

Results

The mean ratings for the experiment are presented in Table 1. Analysis of variance comparisons reveal that the missing VP_2 was rated the best of the three missing VP conditions (compared with missing VP_1: $F1(1,39) = 21.7$, $P < .001$; $F2(1,11) = 16.5$, $P < .005$; compared with missing VP_3: $F1(1,39) = 10.2$, $P < .005$; $F2(1,11) = 12.3$, $P < .005$). There was no difference between the missing VP_1 and the missing VP_3 conditions, nor was there a difference between the missing VP_2 and the all-three VPs conditions ($Fs < 1$).

TABLE 1
Complexity Ratings and (Standard Errors) for the Structures Tested in the Experiment.
Higher Ratings Reflect Higher Complexity

	All three VPs	Missing VP₁	Missing VP₂	Missing VP₃
Rating	2.90 (.12)	3.58 (.14)	2.97 (.14)	3.53 (.14)

Discussion

The results confirm that sentence structures like (3) and (4) are as acceptable as the grammatical versions, despite the fact that they are ungrammatical, missing an obligatorily required VP. Furthermore, the results confirm Fodor's intuition that it is the second VP which is omitted in the acceptable interpretation of these ungrammatical structures. The results disconfirm the disappearing syntactic nodes hypothesis, which applies to delete structures only in RCs with nonlexical RC pronouns. This hypothesis predicts that all three conditions which are missing a VP should be equally bad, and worse than the grammatical version, a prediction which was disconfirmed. The results are also not consistent with the least recent nodes hypothesis, because it is the middle VP, not the first VP, which is forgotten. The results can be accounted for by the high memory cost pruning hypothesis within the SPLT syntactic complexity framework, because the predictions associated with the first RC (the middle VP) require the most memory resources according to this theory, with the consequence that they would be the predictions that would be forgotten, resulting in the missing VP effect.

One aspect of Fodor's intuition regarding the comparison between the ungrammatical but acceptable missing VP structure and its grammatical counterpart that was not confirmed is the intuition that the structure with only two VPs is *less* complex than the one with all three VPs included. No difference was observed in the results reported here for the two comparable conditions. A possible cause of the lack of an observed difference between the grammatical three VP examples and the missing VP_2 examples is the off-line nature of the task. In particular, although there were many fillers in the experiment, participants might still have noticed the multiply embedded structures, and sometimes have looked back to see if the initial NPs were licensed by the verbs that occurred. Although participants were asked to read each sentence only once, re-reading would have been possible in a questionnaire experiment, where the complete sentences are always there for the participant to re-examine. Evidence in favour of this interpretation of the data is provided by Christiansen (1997) in a recent follow-up to the present study using an on-line self-paced word-by-word (centre-presentation) grammaticality judgement task on half of the items used here, in addition to other items. At each point in reading the sentences, participants were asked to judge whether the input string was grammatical so far. At the end of each sentence, participants rated the sentence on a 7-point scale. In self-paced reading, participants cannot re-read words previously analysed, so the strategy of looking back for NPs, which are not licensed by a verb, cannot be applied. Christiansen found that the missing VP_2 condition was

perceived as significantly more grammatical than the grammatical condition with all three verbs, thus confirming the remaining component of Fodor's intuition.

GENERAL DISCUSSION

The results of this experiment confirm the existence of a missing VP effect and disconfirm two early proposed theories of this effect, the disappearing syntactic nodes hypothesis (Frazier, 1985) and the least recent nodes hypothesis (Gibson, 1991). The results are consistent with the third theory considered here, the high memory cost pruning hypothesis within the SPLT syntactic complexity framework. Although compatible with this theory, there are other possible interpretations of the missing VP effect. For example, Christiansen (1997) and Christiansen and Chater (in press) have implemented connectionist models that simulate these and some related nesting results using simple recurrent networks (SRNs, Elman, 1990, 1991). However, the connectionist models are not currently well understood, with the consequence that it is difficult to know what their predictions are with respect to other constructions not yet tested (e.g. the complexity difference between embedding an RC within a sentential complement and the reverse embedding, Cowper, 1976; Gibson, 1991, 1998; Gibson & Thomas, 1998) without running the system on the constructions in question. One possible way of interpreting the connectionist models is in terms of resources. Connectionist networks have resources, in the form of hidden units performing the prediction of the next word in the input string. Limitations in the size of this set of units will limit the network's ability to perform the task. Theories like the SPLT may offer a different level of explanation of these resource limitations.[3] Thus, although Christiansen and Chater's results are intriguing, it is also worth pursuing other, higher-level, accounts of the missing VP effect. We may later be able to reconcile these higher-level accounts with connectionist models like that of Christiansen and Chater's (see also Tabor, Juliano, & Tanenhaus' 1997 connectionist model of ambiguity resolution).

An important issue not yet discussed that the missing VP effect can potentially address is what the units of representation are that are being forgotten. This should indirectly tell us something about what kinds of representations are being retained (and subsequently forgotten) in

---

[3] Christiansen and Chater investigate the effect of changing the number of hidden units. Although they downplay the effect, performance clearly improves with a greater number of hidden units. They seem to regard the possibility that the behaviour of their networks might be described in other terms (such as memory limitations) as an objection to their model, although it is not clear why.

sentence comprehension. One possibility, suggested by the SPLT, is that the representations being retained are syntactic expectations. A second possibility is that the word-strings on which the expectations depend are being retained. One prediction made by the syntactic-expectation representation theory is that, if the memory costs are at the limits of the capacity, then it should be possible to forget one of two costly expectations dependent on the same word. The word-string representation theory predicts that such a situation should not be possible: Both expectations should be forgotten, because they are dependent on the same lexical element. Intuitive judgement evidence suggests that the word-string representation theory makes the correct prediction. In particular, recall that the maximal memory cost in processing (1) occurs at the point of processing "the clinic", at which point the memory load is $2M(2) + 2M(1)$ MUs:

(1) # The patient who the nurse who the clinic had hired admitted met Jack.

There are two syntactic expectations with equivalent maximal memory cost at this point: the expected verb and empty-NP associated with the first RC, both of which depend on the first RC pronoun "who". Forgetting one of these expectations would be enough to bring the total memory load to $M(2) + 2M(1)$ MUs. This resultant memory cost would be below the capacity, as evidenced by the acceptability of a construction consisting of an RC embedded within a sentential complement as in (16) (Cowper, 1976; Gibson, 1991; Gibson & Thomas, 1998):

(16) The fact that the employee who the manager hired stole office supplies worried the executive.

The maximal memory complexity of this construction occurs at the point of processing the NP "the manager", at which point the memory load is $M(2) + 2M(1)$ MUs corresponding to: (a) the expectation for the matrix verb; (b) the expectation for a verb for the sentential complement of "fact" [$M(2)$ MUs]; (c) the expectation for a verb for the RC initiated by "who" [$M(1)$ MUs]; and (d) the expectation for an empty-category coindexed with "who" within the RC [$M(1)$ MUs]. Despite this relatively high memory cost, this construction is quite processable. Thus, forgetting one of the RC expectations in (1) at its point of maximal complexity would be enough to bring the total memory load below the capacity.

But it appears that the processor does not selectively forget one of these expectations: both are forgotten, as evidenced by the acceptability of (3) (the missing VP effect) and the unacceptability of both (17a) [repeated from (6)], in which only the required empty-NP is omitted, and (17b), in which only the required verb is omitted:

(3) $\checkmark$* The patient who$_i$ the nurse who$_j$ the clinic had hired e_j met Jack.

(17)a. * The patient who$_i$ the nurse who$_j$ the clinic had hired e_j was smiling met Jack.

   b. * The patient who$_i$ the nurse who$_j$ the clinic had hired e_j for e_i met Jack.

If only the expected empty-NP for the outer RC were forgotten, then it should be possible to complete the sentence with a verb and an empty-NP for the inner RC, a verb for the embedded RC, and a matrix verb, as in (17a). But this completion is clearly unacceptable. Similarly, omitting the outer RC verb but including an empty-NP site following a preposition also results in an unacceptable structure as in (17a). Both of these continuations are much less acceptable than (3) in which both are omitted. These observations therefore cast doubt on the syntactic-expectation representation theory. They are compatible with the word-string representation theory, according to which the word-strings associated with high memory costs are forgotten. (Note that it is not possible to determine whether the representations implicit in Christiansen and Chater's network model are word-string-based or syntactic-expectation-based, or are based on some other representation type, without testing the kinds of examples discussed here in their model.)

Given the word-string representation theory together with a linguistic complexity theory like the SPLT, other forgetting theories can be suggested. For example, it has long been observed that people's memory for lists of items has a recency component, whereby the most recent items in the list are recalled well, and a primacy component, whereby the earliest items in the list are recalled well (see e.g. Anderson, 1994; Baddeley, 1990; Lewis, 1996, for recent summaries). If people's memory for linguistic structures works in a similar fashion, one would expect similar recency and primacy effects in remembering the word-strings associated with syntactic expectations. The recency/primacy theory makes the same prediction as the high memory cost pruning hypothesis with respect to the doubly nested RC examples explored in this paper. In particular, the three word-strings associated with syntactic expectations at the point of processing the most embedded subject in a doubly nested RC structure are: (i) the first subject; (ii) the first RC pronoun; and (iii) the second RC pronoun. A recency/primacy theory of short-term memory would naturally predict that the middle item is forgotten, the first RC pronoun, resulting in the missing VP effect.

In addition to accounting for the missing VP effect, the recency/primacy account might potentially account for the attachment preferences involving three potential NP sites. In constructions such as (18), reading time studies have shown that attachment to the most local site is easiest

(18a), followed by attachment to the least local site (18c), with attachment to the middle site (18b) hardest of the three (see Gibson, Pearlmutter, Canseco-Gonzalez, & Hickok, 1996, for English and Spanish evidence; see Hemforth, Konieczny, & Scheepers, in press, for German evidence):

(18)a. Attachment of RC to the most local site:
   [$_{NP_1}$ the lamps near [$_{NP_2}$ the paintings of [$_{NP_3}$ the house]]] [$_{RC}$ that was damaged in the flood]

   b. Attachment of RC to the middle site:
   [$_{NP_1}$ the lamps near [$_{NP_2}$ the painting of [$_{NP_3}$ the houses]]] [$_{RC}$ that was damaged in the flood]

   c. Attachment of RC to the first site:
   [$_{NP_1}$ the lamp near [$_{NP_2}$ the paintings of [$_{NP_3}$ the houses]]] [$_{RC}$ that was damaged in the flood]

The extreme difficulty associated with attaching an RC to the middle site might follow from a recency/primacy account, because it may be hard to hold all three attachment sites in memory, especially as none requires any dependent to follow. Consequently, the middle site would be the first to be forgotten.[4]

A recent experimental result relevant to distinguishing the high memory cost pruning hypothesis from the recency/primacy account is provided by Kaan & Gibson (1998) who used an off-line acceptability questionnaire similar to the one reported here to evaluate whether the outer RC VP could be acceptably omitted from a doubly nested RC structure following the matrix verb, as in (19):

(19) Post-verbal doubly nested RC structure:
   The news broadcast supported the labor union which the workers who the corporation was firing (had infuriated).

---

[4] Gibson et al. (1996) consider a recency/primacy account of the three-NP-site attachment data but do not follow it up for two reasons. First, there are cross-linguistic differences in attachment preferences in two-NP-site attachment preferences (Cuetos & Mitchell, 1988). Assuming that the same factors are responsible for attachment preferences when there are two potential sites as when there are three, recency and primacy should also determine two-NP-site attachment preferences. But there is no reason why either of these factors should vary depending on the language, so it is difficult to see how this approach can account for the cross-linguistic differences. However, it is possible that there are other factors in addition to recency and primacy which are responsible for the cross-linguistic differences.

Second, although there is a nonlocal attachment factor affecting attachments to NPs (e.g. the primacy preference), there does not seem to be such a factor in attachments to clauses. Rather, there seems to be a straight locality preference in clauses (Gibson, Pearlmutter, & Torrens, in press). However, this observation may not be a problem for the primacy account of NP attachment preferences, if clauses are treated differently by the processing mechanism than NPs, as some studies have shown (e.g. Roberts et al., 1997, and the references there).

(20) Pre-verbal doubly nested RC structure:
The labor union which the workers who the corporation was firing (had infuriated) was supported by the news broadcast.

In the pre-matrix-verb RC construction in (20), the structures with the middle VP omitted were rated as slightly better than the structures with all three VPs present, similar to the results reported by Christiansen (1997). In contrast, in the post-matrix-verb RC construction in (19), the structures with a missing VP were rated as significantly worse than the structures with both required VPs. At first glance, these results seem to fit the recency/primacy theory better than the high memory cost pruning hypothesis. In pre-verbal contexts, there are three word-strings associated with obligatory syntactic expectations: the matrix subject NP and the two RC pronouns. In post-verbal contexts, there are only two: the two RC pronouns. With only two word-strings with obligatory syntactic expectations, a recency/primacy account predicts that neither should be forgotten, even in a high memory load situation, because one is favoured by recency and the other by primacy. Thus, the poor ratings of the post-verbal missing VP items is consistent with this theory.

On the other hand, the high memory cost pruning hypothesis predicts that the expectations associated with the first RC pronoun should be forgotten, because of the high SPLT memory load of the structure. This prediction was not confirmed. However, other aspects of the results suggest a slight variant of the SPLT memory load hypothesis which is consistent with the results. In particular, the post-verbal RCs were rated slightly better than the pre-verbal RCs in Kaan and Gibson's experiment, contrary to earlier results obtained by Gibson and Thomas (1997) on similar constructions. To account for these results within the SPLT framework, it is possible that the prediction of the top-level predicate is not zero-cost as assumed by Gibson (1998), but rather is associated with some constant (nonincreasing) quantity of memory cost. As a result, post-verbal structures should be slightly less complex than pre-verbal ones. If the post-verbal RCs are indeed less complex, then the maximal memory load incurred during the processing of this structure might be within the capacity of the system, with the consequence that omitting a VP from this construction would make it obviously ungrammatical, as compared with its control containing both required VPs, which is complex but processable. Further experiments are necessary to distinguish this hypothesis from the recency/primacy account and other possible accounts.

Concluding remarks

This paper has reported an experiment that provides a surprising result: That sentence structures with a missing grammatically required constituent

sound as good as or better than the structures with all grammatically required constituents. This effect was interpreted in terms of the processor forgetting parts of its initial representation of an input string at points of high memory complexity. The observed acceptability patterns suggest that the representations that are retained (and subsequently forgotten) in processing sentences consist of the lexical word-strings processed thus far, and not the syntactic expectations associated with these word-strings. A number of specific accounts of forgetting were considered, three that warrant further investigation: (1) the high memory cost pruning hypothesis; (2) a recency/primacy account; and (3) a connectionist account, which might possibly be a different level of explanation of one of the two preceding accounts, or which could have other principles underlying it that are not yet well understood.

REFERENCES

Anderson, J.R. (1994). *Learning and memory.* Wiley, New York.

Baddeley, A. (1990). *Human memory.* Hove, UK: Lawrence Erlbaum Associates Ltd.

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

Bever, T.G., & Townsend, D. (1979). Perceptual mechanisms and formal properties of main and subordinate clauses. In W.E. Cooper & E.C.T. Walker (Eds), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 159–226). Hillsdale, NJ: Erlbaum.

Bresnan, J. (1982). *The mental representation of grammatical relations.* MIT Press, Cambridge, MA.

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

Chomsky, N., & Miller, G.A. (1963). Introduction to the formal analysis of natural languages. In R.D. Luce, R.R. Bush, & E. Galanter (Eds), *Handbook of mathematical psychology* (Vol. 2, pp. 269–321). Wiley, New York.

Christiansen, M.H. (1997, March). *The (un)grammaticality of double center-embedded sentences: A connectionist perspective.* Poster presented at the 10th CUNY Sentence Processing Conference, Santa Monica, CA.

Christiansen, M.H., & Chater, N. (in press). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science.*

Church, K.W. (1980). *On memory limitations in natural language processing.* Unpublished master's thesis, Massachusetts Institute of Technology, Cambridge, MA.

Cowper, E.A. (1976). *Constraints on sentence complexity: A model for syntactic processing.* Ph.D. dissertation, Brown University, Providence, RI.

Cuetos, F., & Mitchell, D.C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition, 30,* 73–105.

Elman, J.L. (1990). Finding structure in time. *Cognitive Science, 14*, 179–211.

Elman, J.L. (1991). Distributed representations, simple recurrent networks and grammatical structure. *Machine Learning, 7*, 195–225.

Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies*. Unpublished doctoral dissertation, University of Connecticut, Storrs, CT, USA.

Frazier, L. (1985). Syntactic complexity. In D. Dowty, L. Karttunen, & A. Zwicky (Eds), *Natural language processing: Psychological, computational and theoretical perspectives* (pp. 129–189). Cambridge, UK: Cambridge University Press.

Frazier, L., & Fodor, J.D. (1978). The Sausage Machine: A new two-stage parsing model. *Cognition, 6*, 291–325.

Garrod, S., & Sanford, A.J. (1977). Interpreting anaphoric relations: The integration of semantic information while reading. *Journal of Verbal Learning and Verbal Behavior, 16*, 77–90.

Garrod, S., & Sanford, A.J. (1982). The mental representation of discourse in a focused memory system: Implications for the interpretation of anaphoric noun phrases. *Journal of Semantics, 1*, 21–41.

Garrod, S.C., & Sanford, A.J. (1994). Resolving sentences in a discourse context: How discourse representation affects language understanding. In Gernsbacher, M.A. (Ed.), *Handbook of psycholinguistics* (pp. 675–698). New York: Academic Press.

Gibson, E. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA.

Gibson, E. (1998). Syntactic complexity: Locality of syntactic dependencies. *Cognition, 68*, 1–76.

Gibson, E., Pearlmutter, N., Canseco-Gonzales, E., & Hickok, G. (1996). Recency preference in the human sentence processing mechanism. *Cognition, 59*, 23–59.

Gibson, E., Pearlmutter N., & Torrens, V. (in press). Recency and lexical preferences in Spanish. *Memory and Cognition*.

Gibson, E., & Thomas, J. (1998). *The complexity of English nested structures*. Unpublished manuscript, Massachusetts Institute of Technology, Cambridge, MA.

Hakes, B., Evans, J., & Brannon, L. (1976). Understanding sentences with relative clauses. *Memory and Cognition, 4*, 283–296.

Haliday, M.A.K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Haviland, S.E., & Clark, H.H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior, 13*, 512–521.

Hawkins, J.A. (1994). *A performance theory of order and constituency*. Cambridge, UK: Cambridge University Press.

Hemforth, B., Konieczny, L., & Scheepers, C. (in press). Syntactic attachment and anaphor resolution: Two sides of relative clause attachment. In M. Crocker, M. Pickering, & C. Clifton, Jr. (Eds), *Architectures and mechanisms for language processing*. Cambridge, UK: Cambridge University Press.

Jarvella, R.J. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior, 10*, 409–416.

Jarvella, R.J. (1979). Immediate memory and discourse processing. In G.H. Bower (Ed.) *The psychology of learning and motivation* (Vol. 13, pp. 379–421). New York: Academic Press.

Joshi, A.K. (1990). Processing crossed and nested dependencies: An automaton perspective on the psycholinguistic results. *Language and Cognitive Processes, 5*, 1–27.

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

Kaan, E., & Gibson, E. (1998). *Structural forgetting in complex linguistic structures*. Cambridge, MA: Department of Brain and Cognitive Sciences: Massachusetts Institute of Technology. (In preparation.)

Kac, M.B. (1981). Center-embedding revisited. In *Proceedings of the third annual conference of the Cognitive Science Society* (pp. 123–124). Hillsdale, NJ: Erlbaum.

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, *2*, 15–47.

Lewis, R. (1993). *An architecturally-based theory of human sentence comprehension*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA.

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

Marslen-Wilson, W., Tyler, L.K., & Seidenberg, M. (1978). Sentence processing and the clause boundary. In W.J.M. Levelt & G.B. Flores d'Arcais (Eds), *Studies in the perception of language* (pp. 219–246). New York: Wiley.

Miller, G.A., & Chomsky, N. (1963). Finitary models of language users. In R.D. Luce, R.R. Bush, & E. Galanter (Eds), *Handbook of mathematical psychology* (Vol. 2, pp. 419–491). New York: Wiley.

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

Milsark, G.L. (1983). On length and structure in sentence parsing. *Cognition, 13*, 129–134.

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

Rambow, O., & Joshi, A. (1994). A processing model for free word-order languages. In C. Clifton, L. Frazier, & K. Rayner (Eds), *Perspectives in sentence processing* (pp. 267–301). Hillsdale, NJ: Erlbaum.

Roberts, R., Hooker, C., & Gibson, E. (1997, March). *Sentence memory capacity and working memory*. Poster presented at the 10th CUNY Sentence Processing Conference, Santa Monica, CA.

Stabler, E.P. (1994). The finite connectivity of linguistic structures. In C. Clifton, L. Frazier, & K. Rayner (Eds), *Perspectives in sentence processing* (pp. 303–336). Hillsdale, NJ: Erlbaum.

Tabor, W., Juliano, C., & Tanenhaus, M.K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes, 12*, 211–272.

Warren, T., & Gibson, E. (1998). *Discourse accessibility and linguistic complexity*. Unpublished manuscript, Cambridge, MA: Massachusetts Institute of Technology.

Waugh, N.C., & Norman, D.A. (1965). Primary memory. *Psychological Review*, *72*, 89–104.

APPENDIX

Experimental items

All four conditions are given for the first item. Only the condition with all three verb phrases is given for the remaining items.

1. All three VPs: The ancient manuscript that the grad student who the new card catalog had confused a great deal was studying in the library was missing a page.

   Missing VP_1: The ancient manuscript that the grad student who the new card catalog was studying in the library was missing a page.

Missing VP$_2$: The ancient manuscript that the grad student who the new card catalog had confused a great deal was missing a page.

Missing VP$_3$: The ancient manuscript that the grad student who the new card catalog had confused a great deal was studying in the library.

2. The lullaby that the famous country singer who the record label had signed to a big contract was singing yesterday was written seventy years ago.

3. The game that the child who the lawnmower had startled in the yard was playing in the morning lasted for hours.

4. The crime that the gangster who the story had profiled had planned for weeks was quickly solved.

5. The picture that the artist who the school had expelled for cheating was hurriedly copying was printed in a magazine.

6. The trophy that the athlete who the restaurant had hired as a spokesman had won at the track meet was stolen later.

7. The apartment that the maid who the service had sent over was cleaning every week was well decorated.

8. The shirt that the seamstress who the immigration officer had investigated last week was carefully mending needed to be dry cleaned.

9. The lecture that the professor who the newspaper story had just profiled in detail was teaching poorly was not well attended.

10. The novel that the horror author who the publishing company had recently fired had typed quickly was banned by the local library.

11. The prayer that the monk who the religious fanatic had persecuted relentlessly was chanting every day was echoing in the empty church.

12. The monologue that the actor who the movie industry had snubbed repeatedly was performing last month was extremely well written.