The Role of Experience in Sentence Acceptability and Sentence Processing

Dave Braze

Haskins Laboratories
braze@haskins.yale.edu

The 16th Annual CUNY Conference on Human Sentence Processing

Cambridge, MA
March 27-29, 2003

Introduction

This study provides evidence from the phenomenon of 'syntactic satiation' that some syntactic islands may arise from processing limitations rather than grammatical proscriptions. Syntactic satiation is a phenomenon where some (but not all) unacceptable sentence types become increasingly acceptable with repeated exposure. The eye-movement study reported here uncovers a reading time phenomenon that co-varies with judgment satiation.

Using an acceptability judgment task, Snyder (2000) showed that satiation occurs with very few exposures to some sentence types, overlaping suggestively with 'weak' island violations. He also argued that no simple mapping exists between degree of initial acceptability of a type and its susceptibility to satiation. Of the seven sentence types evaluated in Snyder's judgment satiation experiment, two were selected for inclusion in the present eye-movement study. The first, argument extraction from *whether* clauses, was selected because it had shown the clearest satiation effect; the second, argument extraction from *while* adjuncts, because it had proved completely resistant to satiation. Additionally, the current study contrasted a felicitous baseline condition with each island-violation type and coherent contexts were provided for both island violations and baselines.

The Present Study

Of central interest is whether an eye-movement analog to judgment satiation exists. Or, more directly, we can ask the question of whether satiation effects are task-specific. In order to answer this question, I examine whether the processing effort expended on *whether*-islands and adjunct-islands changes with repeated exposure. Processing difficulty is gauged at each repetition by comparison of reading times on island-violations with matched control sentences.

An analog to judgment satiation should surface as a change in the reading times on *whether*-island violations relative to their matched control sentences, as readers encounter additional tokens of each type. Conversely, reading times on adjunct-island violations should not change with repeated exposure, relative to their matched controls.

There are two parts to the present study. First, an acceptability judgment task established that materials developed for this project were adequate to evoke satiation. Next, a different participant group read the same sentences while their eye-movements were recorded. Here, the task was to read for comprehension; no acceptability judgments were required. Comprehension probes ensured that participants attended to the task of reading for understanding. The acceptability judgment task and the eye-movement task used the same materials.

Experimental materials

The present study incorporates the two sentence types evaluated in Snyder (2000) which showed the greatest contrast in satiation: argument extraction from *whether*-islands and from adjunct-islands. Twenty-eight sets of sentences of each type were devised as shown in (1) and (2), for a total of fifty-six test items. Each set consisted of a declarative context sentence, an island violation and a grammatical control sentence.

(1) a. After the meeting, Valerie wondered whether Don had liked the story.
b. When did Valerie wonder whether Don had liked the story?
c. What did Valerie wonder whether Don had liked?

(2) a. On Saturday, Fraser cleaned the bathroom while Judy mopped the kitchen.
b. When did Fraser clean the bathroom while Judy mopped the kitchen?
c. What did Fraser clean the bathroom while Judy mopped on Saturday?

Initially, two stimulus lists were constructed based on these materials. Each list contained one sentence from each matched pair for a total of fifty-six test items per list. Thus, each list contained an equal number (14 each) of *whether*-island violations, *whether*-island controls, adjunct-island violations and adjunct-island controls.

Critical items were interleaved among a like number of foils. Foils were designed to offset the fact that, among critical items, all island violations began with the word *what* while all control sentences began with *when*. So, one-half of foils included grammatical questions beginning with *what*, and one-half contained ungrammatical questions beginning with *when*, as shown in (3) and (4). In the eye-tracker task, grammatical foils included comprehension probes.

Within lists, items were arranged in blocks to ensure that sentence types were evenly distributed throughout. Each of fourteen blocks contained eight sentences: one *whether*-island violation, one *whether*-island control, one adjunct-island violation and one adjunct-island control, as well as four foils (two of which were ungrammatical); the order of item types within each block was randomized. Finally, two additional stimulus lists were created by reversing the block order of the original lists; forward/reverse block order was counterbalanced across subjects.

(3) context: Yesterday, Dan said Laura told him she'd seen bigfoot in her yard.
question: What did Dan say Laura told him?
probe: That she'd seen bigfoot in her yard. (TRUE)

(4) context: Sam doesn't know that, last week, Joey fixed the car.
question: When doesn't Sam know that Joey fixed the car?

These four lists were used first in a judgment satiation task to ascertain that materials were adequate to elicit judgment satiation parallel to that reported in Snyder (2000). After vetting the materials, the same lists were used in an eye-tracker study in which participants' only task was to read for comprehension. No acceptability judgments were required.

Judgment Satiation

Method

Subjects

Forty-two undergraduate students received course credit for participation. All were native speakers of English. Participants had no knowledge of the purpose of the experiment and no prior exposure to the test materials.

Procedure

In small groups, participants read context/question pairs from computer monitors. They were instructed to read each pair and then to make a YES/NO acceptability judgment for the question. Comprehension probes were *not* included in the judgment task.

Note that, in the lists used for the judgment task, a counter-balancing error resulted in two presentation lists containing 9 adjunct-island violations and 19 *whether*-island violations (rather than 14 of each), while the other two lists contained 19 adjunct-island violations and 9 *whether*-island violations. Therefore, only the first 9 of each sentence type were included in the judgment task analysis. The error was corrected before carrying out the eye-movement component of the study.

Results and Discussion

Judgment satiation effects were evaluated using a modified sign test in the same manner as Snyder (2000). Results indicate a clear satiation effect for *whether*-island violations ($p<.005$), but no such effect for adjunct island violations. This replicates previous findings and demonstrates that sentence materials designed for use in the present study are qualitatively similar to those used in studies of judgment satiation.

Satiation in Reading Times

Method

The same materials used in the judgment satiation task were used in the eye-movement task.

Subjects

Twenty-eight undergraduate students were paid to participate. All were native speakers of English with vision they reported to be normal or corrected to normal with contact lenses. None of the participants in the judgment satiation task were included in the eye-tracker group. Participants had no knowledge of the purpose of the experiment and no prior exposure to the test materials.

Apparatus

Eye-movements were recorded with a Skalar 6500 eye-tracker. The system uses infra-red transducers positioned in front of the eye to detect eye-movements. The output is a continuously varying analog signal corresponding to gaze direction, which was digitized with a 1 ms sampling rate. A forehead rest and individually prepared bite bars were used to stabilize participants' head positions in order to minimize noise in the signal due to head movements. Sentences were presented on a computer monitor positioned 64 centimeters from the participants' eyes.

Procedure

Sentences were viewed one at a time. Each context sentence appeared alone on the screen on a single line. Participants read these sentences and then clicked a button to clear the context sentence and display the associated *wh*-question. The question also appeared on a single line on the screen. After one-half of foils, participants were presented with a possible answer to the question and required to indicate whether the answer was TRUE or FALSE. True and False probes occurred with equal probability. The purpose of these probes was simply to ensure that participants were reading for comprehension. No participant missed more than four of the twenty-eight comprehension probes; response accuracy does not enter into the analyses reported below.

For purposes of analyzing the eye-movement data, sentences were divided into regions as shown in (5) and (6). The dependent measure we focus on here is first-pass regional reading time.

(5) a. When did●Valerie●wonder●whether●Don●had liked●the story?
b. What did●Valerie●wonder●whether●Don●had liked?

(6) a. What did●Fraser●clean●the bathroom●while●Judy mopped●on Saturday?
b. When did●Fraser●clean●the bathroom●while●Judy mopped●the kitchen?

Results

For each island type, a mixed-models analysis of covariance for generalization to subjects, incorporating the factor island-violation (island violation, grammatical control) and the covariate repetition was carried out at each sentence region. Main effects of island-violation are in evidence for both *whether*-islands and adjunct islands.

In adjunct-islands, a main effect of the island-violation on reading times is apparent. **Figure 1** shows reliable reading time differences at the object of the main clause (region 4) [$F1(1,25)=5.19$, $p<.05$, $MSE=18475$] and *while* (region 5) [$F1(1,25)=12.16$, $p<.01$, $MSE=6620$]. At sentence region 6, reading times for island violations are numerically greater than reading times for controls, but the difference is not reliable.

As shown in **Figure 2**, the main effect of whether-island violations first surfaces at the main verb of the matrix clause, *wonder*, in sentence region 3 [$F1(1,25)=9.27$, $p<.01$, $MSE=4046$], but the modal effect occurs at the subject of the embedded clause, region 5 [$F1(1,25)=13.24$, $p=.001$, $MSE=12391$]. The effect persists to region 6, which contains the verb of the embedded clause [$F1(1,25)=4.45$, $p<.05$, $MSE=95589$]. However, differences between *whether*-island violations and their controls that surface in sentence region 6 must be interpreted with caution due the fact that *whether*-island violation sentences end at region 6, while control sentences continue for another two to three words.

Figure 4 shows a clear analog to judgment satiation in first-pass regional reading time attributable to *whether*-island violations. A significant interaction of island condition and repetition in sentence region 5 [$F1(1,508)=7.63$, $p<.01$, $MSE=14500$] suggests that the effort expended in processing *whether*-island violations, relative to control sentences, increases with repeated exposure to sentences of that type. Sentence region 5 is the point at which the modal reading time effect for whether-island violations occurs. A similar interaction is present in region 4 [$F1(1,537)=7.43$, $p<.01$, $MSE=7012$].

Conversely, **Figure 3** exemplifies the fact that perturbations due to adjunct island violations, relative to their controls, do not change over the course of a session. Crucially, separate ANCOVAs were carried out at each sentence region, using reading times and incidence of regressive eye-movements as dependent measures. In the case of adjunct-islands, **no** sentence region shows an interaction between island condition and repetition on **either** dependent measure.

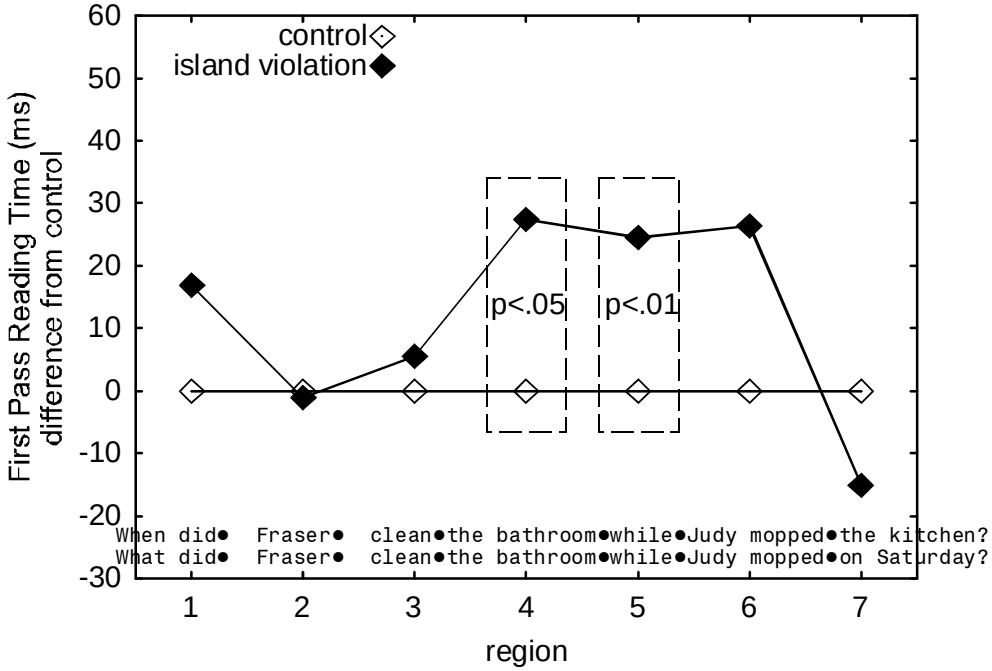

Figure 1: Adjunct-island violations – regional first-pass reading times plotted as differences from the non-island violation control condition.

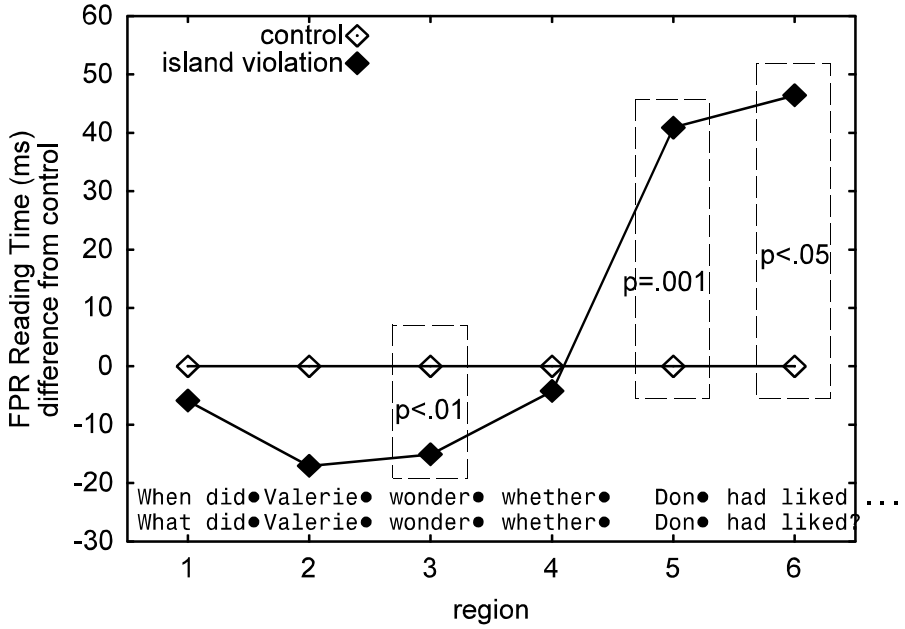

Figure 2: *Whether*-island violations – mean regional first pass reading times plotted as difference from non-island violation controls.

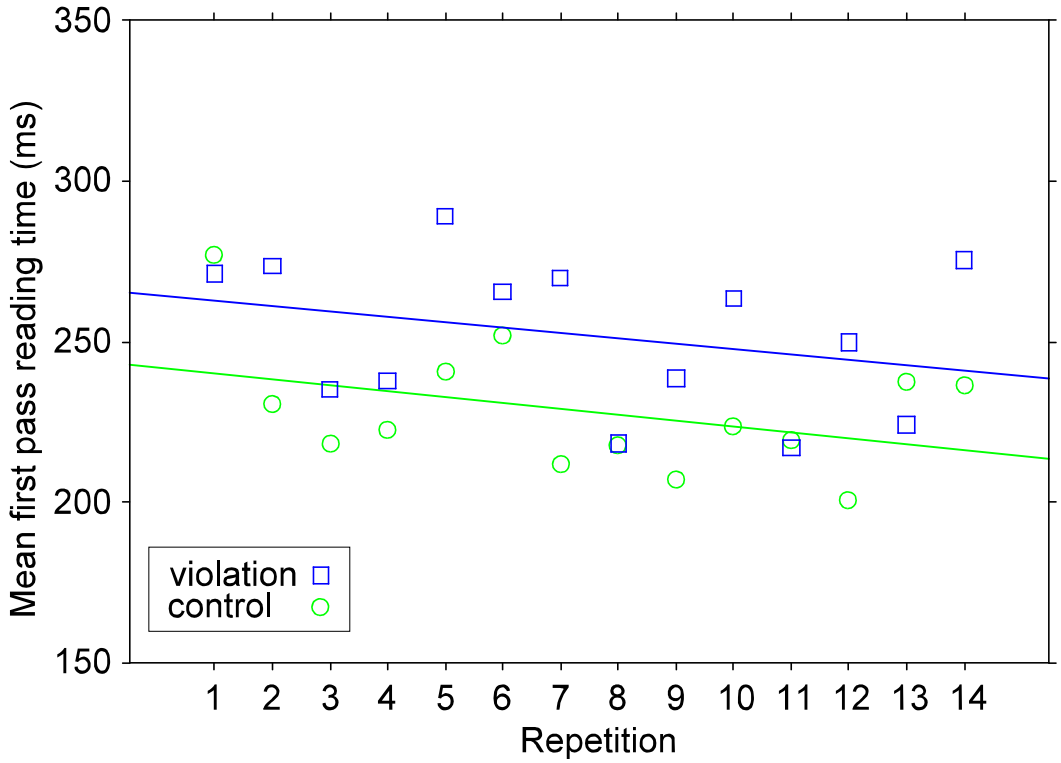

Figure 3: Adjunct-islands – mean first pass reading times at region 5 for island violations and controls for each of 14 repetitions, with linear fits to each type.

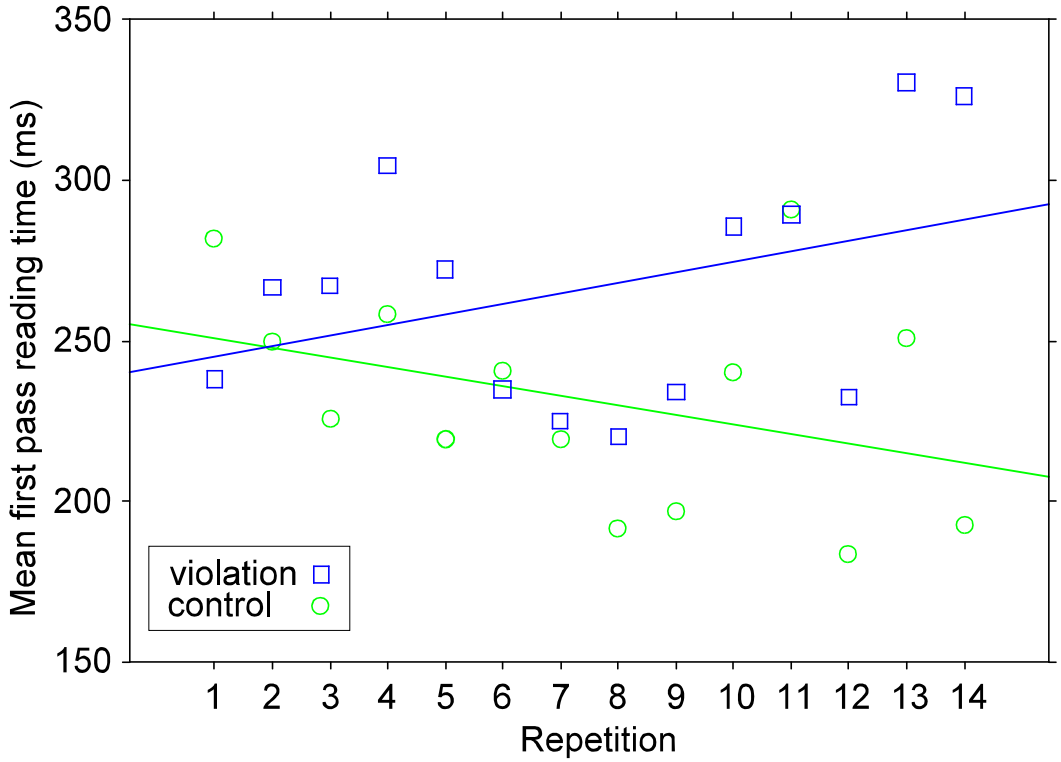

Figure 4: *Whether*-islands – mean first pass reading times at region 5 for island violations and controls for each of 14 repetitions, with linear fits to each type.

Discussion

Both island types show elevated reading times relative to controls. These main effects of island condition are consistent with previous work indicating early sensitivity of the parsing mechanism to island configurations (Stowe, 1985; McKinnon & Osterhout, 1996; McElree & Griffith, 1998). A demonstration that the parser is sensitive to island configurations is hardly surprising. However, participants responded differently to the two island types under the condition of repeated exposure. Regional reading times for whether-islands increase with repetition while reading times for adjunct-islands remain constant, relative to controls. Given the conventional assumption that elevated reading times reflect increased processing effort, we can conclude that effort expended on *whether*-islands increases with experience, indicating the sentence processor may adapt rapidly to demands of certain complex constructions. No such experience-based change is in evidence for adjunct-island violations.

These results demonstrate an on-line counterpart of judgment satiation in that *whether*-islands induce progressively longer reading times as participants are exposed to repetitions of these structures. This increase corresponds to an increase in acceptability for *whether*-islands with repeated exposure. Conversely, repeated exposure to adjunct-islands brings about neither a change in reading times nor in acceptability.

The contrast between *whether*-island violations and adjunct-island violations suggests that, at a minimum, the intuitions of unacceptability that native English speakers report for these configurations have different etiologies. One especially interesting possibility is that the perceived anomaly of *whether*-island violations derives from a non-grammatical constraint.

Based on the results of a sentence judgment task, Kluender and Kutas (1993) suggested that processing demands associated with crossing a clause boundary, together with the demands of filler-gap association and those related to the processing of words that are in some sense semantically "heavy" (like *if* or *who*), combine to generate a point of exceptional burden for the processor. It is this processing overload that leads to the intuition of unacceptability in sentences like (7b) and (7c). An embedded *if* complement, like (7b), is directly comparable to the *whether*-island clauses used in the present experiment. Each instantiates an embedded yes/no question.

(7) a. What did you figure out THAT you should tell the boss about *e* before the meeting?
b. What can't you figure out IF you should tell the boss about *e* before the meeting?
c. What$_i$ can't you figure out WHO$_j$ e_i should tell the boss about e_j before the meeting?

It would be profitable to make Kluender and Kutas' suggestion about semantic processing load more specific. A classical account of the semantics of yes/no questions (Karttunen, 1977) expresses the denotation of such questions as a set of propositions conveying contextually salient possible answers. Given a sentence like (8), the reference of the embedded question is a set containing the two propositions shown in (9). Thus, the object of Valerie's wondering is the veracity of the propositions in that set. The denotation of an embedded yes/no question is a complex referential entity that must be represented in some form of mental discourse model.

(8) Valerie wondered whether Don had liked the story.

(9) { "Don had liked the story" "Don had not liked the story" }

In order to account for the immediacy of island effects, I assume that the mental model is updated rapidly, on a virtually word-by-word basis (Crain & Steedman, 1985). Further, modifications to the mental model (and all other levels of representation) arise through the operation of automatic, pre-cognitive functions (Fodor, 1983). The automatic or reflexive of these functions reflects the fact that the fundamental operand of the language processing system, speech, is ephemeral in nature and short term storage is limited. In the context of sentence level processing, Ferreira and Henderson (1998) suggest that certain operations of the language processing mechanism are temporally bounded, automatically curtailed if processing time exceeds some time limit (which may be subject to individual variation).

It is certain that the computational loads associated with clause embedding, filler-gap association and the semantics of embedded questions are, in isolation, within the capacity of the sentence processing mechanism. However, in the context of *wh*-extraction from an indirect question, the combination of these loads creates a situation in which the sentence processing mechanism is overburdened. It is this overload that leads to the unacceptability of sentences like (10).

(10) What did Valerie wonder whether Don had liked?

How do satiation effects fit into this picture? I assume a model based on the Referential Model of Crain and Steedman (1985). A simple extension to this model hypothesizes that syntactic and semantic and processors share a common resource pool. Further, a load-balancing function serves to dynamically allocate these shared resources among competing demands. Such a model is shown in **Figure 5.**

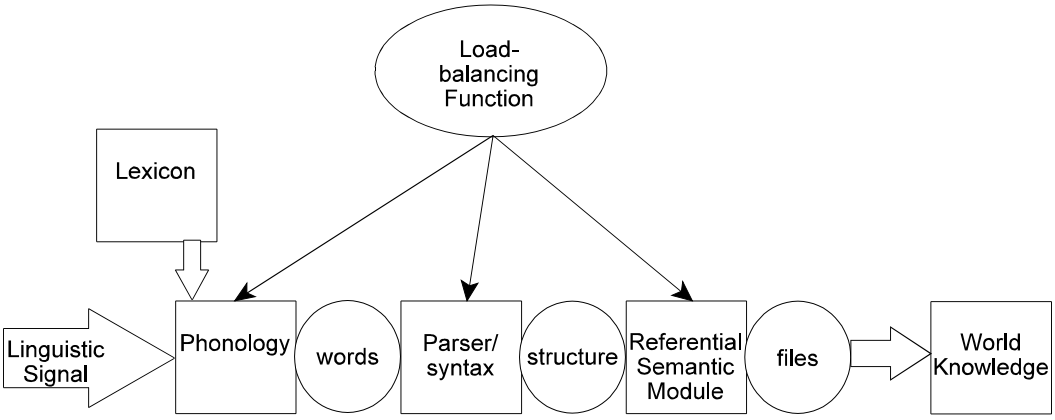

Figure 5: Extended Referential Model.

The load-balancing function is responsible for arbitrating among competing demands of the various sub-processors in order to facilitate the flow of information through the sentence processing system. This function is, by hypothesis, capable of adjusting the allocation of attentional resources to meet the moment-by-moment needs of each sub-processor. I will tentatively identify the load-balancing function as a component of the working memory's central executive function (Baddeley, 1986). When the sentence processing system is confronted with a construction that places heavy demands on more than one sub-processor at the same time (like wh-extraction from *whether*-islands), the load-balancing function attempts to adapt the allocation of attentional resources in such a way as to maximize the effectiveness of the entire system. With repeated exposure to identical constructions,

this "balancing-act" becomes more refined. Thus, as the executive function sees more whether-island violations, it learns to cope with them more efficiently, within the limits of its capacity. This is why the acceptability of whether-island violations increases with repeated exposure.

How does the increase in reading times for *whether*-island violations follow from this proposal? By hypothesis, each sub-processor is allocated some default quantity of the shared resource pool, but the distribution of resources across all sub-processors is sub-optimal for the purpose of comprehending *whether*-island violations. Degraded acceptability in the judgment task occurs when one sub-processor (plausibly, the referential sub-processor, in the case of *whether*-island violations) is unable to fully carry out its function due to the combination of a resource limitation and the inherent temporal boundedness of the system.

By hypothesis, the increasing reading times participants exhibit as a result of repeated exposure to *whether*-island violations arise when the executive function shifts resources to the initially deprived referential sub-processor. This shift in resources allows the referential component to (more) fully carry out its function, taking more time in the process. Thus, the increases in acceptability and in reading time are accounted for by the same mechanism.

Finally, The absence of judgment satiation, or the reading-time cognate, in adjunct-islands may be due to the resource demands of that construction being completely beyond the capacity of the human sentence processing mechanism. Alternately, the underlying cause of the unacceptability of adjunct islands may not lie in resource limitations, but in grammatical constraints.

If this proposal is on the right track, a number of other phenomena, are likely to fall under its scope. Among the more suggestive candidates are *wh*-movement from within the scope of factive predicates and from within the scope of negation and negative verbs.

(11) a. ?? What did Bill confirm that Roger had eaten?
b. What did Bill allege that Roger had eaten?

(12) a. * How well did John deny that he performed?
b. * How well did few critics think that he performed?

Summary

The data reported here support the idea that some syntactic islands may result from processing limitations rather than grammatical constraints. I adopt the suggestion, due to Kluender & Kutas (1993), that at least some syntactic islands arise when processing demands associated with a clause boundary, together with those due to semantically heavy words like *whether*, combine to generate a processing bottleneck. The additional load of filler-gap association across this juncture accounts for the relative unacceptability of *whether*-islands. In order to account for the on-line satiation effect, I proposed an automatic resource management function, which actively shifts resources among competing demands to minimize such bottlenecks. Experience with a demanding construction facilitates optimal (construction specific) resource allocation for subsequent tokens of that construction. This resource management function accounts for the processor's adaptability in the

face of repeated exposure to *whether*-islands. The function is tentatively identified as part of the central executive component of verbal working memory. The sentence processor's failure to adapt to adjunct-islands may point toward a grammatical explanation for such structures, or, alternately, to processing demands that are too extreme to accommodate.

References

Baddeley, A. D. 1986. Working memory. Oxford, UK: Oxford University Press.

Crain, S., and M. Steedman. 1985. On not being led up the garden path: the use of context by the psychological parser. In: D.R. Dowty, L. Karttunen and A.M. Zwicky (eds.), Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives, 320-358. Cambridge: Cambridge University Press.

Ferreira, F., and J.M. Henderson. 1998. Syntactic reanalysis, thematic processing, and sentence comprehension. In: J.D. Fodor and F. Ferreira (eds.), Reanalysis in Sentence Processing, 73-100. Dordrecht: Kluwer Academic Publishers.

Fodor, J.A. 1983. The modularity of mind. Cambridge, MA: MIT Press.

Kluender, R.E., and M. Kutas. 1993. Subjacency as a processing phenomenon. Language and Cognitive Processes 8.4, 573-633.

Kluender, R.E. 1998. On the Distinction Between Strong and Weak Islands: a processing perspective. In: P.W. Culicover and L. McNally (eds.), The Limits of Syntax, 241-280. San Diego, CA: Academic Press.

McElree, B., and T. Griffith. 1998. Structural and lexical constraints on filling gaps during sentence comprehension: A time course analysis. Journal of Experimental Psychology: Learning, Memory, and Cognition 24.2, 432-460.

McKinnon, R., and L.E. Osterhout. 1996. Constraints on movement phenomena in sentence processing: evidence from event related brain potentials. Language and Cognitive Processes. 11.5, 495-523.

Snyder, W. 2000. An Experimental Investigation of Syntactic Satiation Effects. Linguistic Inquiry 31, 575-582.

Stowe, L.A. 1985. Parsing WH constructions: Evidence for on-line gap location. Language and Cognitive Processes. 1.3, 227-245.