

- Rooth, M. (1985). *Association with focus*. Unpublished doctoral dissertation, University of Massachusetts, Amherst.
- Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, 1, 75–116.
- Sedivy, J., Carlson, G., Tanenhaus, M., Spivey-Knowlton, M., & Eberhard, K. (1994). The cognitive function of contrast sets in processing focus constructions. In *Working Papers of the IBM Institute for Logic and Linguistics*.
- Sedivy, J., Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Carlson, G. (1995). Using intonationally-marked presuppositional information in on-line language processing: Evidence from eye movements to a visual model. *Proceedings of the 17th Annual Conference of the Cognitive Science Society* (pp. 375–380). Mahwah, NJ: LEA Associates.
- Spivey-Knowlton, M., & Sedivy, J. (1995). Parsing attachment ambiguities with multiple constraints. *Cognition*, 55, 227–267.
- Spivey-Knowlton, M., Sedivy, J., Eberhard, K., & Tanenhaus, M. (1994). Psycholinguistic study of the interaction between language and vision. In *Proceedings of the 12th National Conference on Artificial Intelligence: Workshop on the Integration of Natural Language and Vision Processing*.
- Spivey-Knowlton, M., & Tanenhaus, M. K. (1994). Referential context and syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing*. Hillsdale, NJ: Erlbaum.
- Spivey-Knowlton, M., & Tanenhaus, M. (1995). *Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency within an integration-competition framework*. Manuscript submitted for publication.
- Spivey-Knowlton, M., Tanenhaus, M., Eberhard, K., & Sedivy, J. (1995). Eye-movements accompanying language and action in a visual context: Evidence against modularity. (Eds.), *Proceedings of the 17th Annual Conference of the Cognitive Science Society* (pp. 25–30). Mahwah, NJ: LEA Associates.
- Spivey-Knowlton, M., Trueswell, J., & Tanenhaus, M. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology*, 37, 276–309.
- Steedman, M. (1987). Combinatory grammars and human language processing. In J. L. Garfield (Ed.), *Modularity in knowledge representation and natural language understanding* (pp. 187–205). Cambridge, MA: MIT Press.
- Swinney, D., & Osterhout, L. (1990). Interference generation during auditory language comprehension. *The Psychology of Learning and Motivation*, 25, 17–33.
- Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). The interaction of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.
- Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (in press). Using eye-movements to study spoken language comprehension: Evidence for visually-mediated incremental interpretation. In T. Inui & J. McClelland (Eds.), *Attention & Performance XVI: Integration in Perception and Communication*.
- Tanenhaus, M., & Trueswell, J. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Handbook of cognition and perception: Vol. 11. Speech and Language*. San Diego, CA: Academic Press.
- Taraban, R., & McClelland, J. (1988). Constituent attachment and thematic role expectations. *Journal of Memory and Language*, 27, 597–632.
- Taraban, R., & McClelland, J. (1990). Sentence comprehension: A multiple constraints view. In D. Balota, K. Rayner, & G. Flores d'Arcais (Eds.), *Comprehension processes in reading*. Hillsdale, NJ: Erlbaum.

## Constraint Satisfaction as a Theory of Sentence Processing

Lyn Frazier<sup>1</sup>

Accepted August 11, 1995

*Various problems with the constraint satisfaction model are discussed. It is argued that the empirical evidence presented in support of the model does not concern predictions of the model that diverge from those of depth-first (one analysis at a time) models. Several methodological problems are also noted. As a theory of sentence processing, the model is inadequate. It fails to account for the assignment of local structure, global structure, structure involving discontinuous dependencies, long-distance dependencies, and adjunct phrases. It makes incorrect predictions about the timing of syntactic analysis. Further, because syntactic structure is available only through activation of syntactic projections stored in the lexical entry of words, the model leaves entirely unexplained the myriad psycholinguistic findings demonstrating independence of lexical and syntactic structure (in Event Related Potential studies, code-switching, pure syntactic priming, etc). Finally, the model is not restrictive or explanatory, providing an account that largely consists of post hoc correlations between frequency counts or subjects' ratings of sentences and processing time data for the same sentences.*

Several recent studies have been presented as support for a competition theory of sentence processing where alternative syntactic analyses are activated simultaneously, with activation levels modulated by frequency and probabilistic contextual constraints (Garnsey, Pearlmutter, Myers, & MacDonald, 1995; MacDonald, Pearlmutter, & Seidenberg, 1994; Pearlmutter, & MacDonald, 1995; Trueswell & Tanenhaus, 1994; Trueswell, Tanenhaus,

The present paper is an excerpt from "Sentence (Re-)Analysis" presented at the 1994 CUNY Conference. It was supported by NIH Grant HD18708 to Clifton and Frazier and DBS9121375 to Rayner and Frazier. I am very grateful to Chuck Clifton and to an anonymous reviewer for comments on an earlier draft of this paper.

<sup>1</sup> University of Massachusetts, Amherst, Massachusetts 01003.

<sup>2</sup> Address all correspondence to Lyn Frazier, Department of Linguistics, University of Massachusetts, Amherst, Massachusetts 01003.

& Garnsey 1994). In this constraint-satisfaction approach, the frequency of alternative lexical and syntactic analyses and probabilistic contextual constraints are claimed to interact continuously to resolve structural ambiguities using the same principles (frequency, sensitivity to context) believed to govern lexical ambiguity resolution. Beyond that, the model is not really specified: The model does not indicate how constituents are identified or how coherent parts (words or phrases) are put together to identify novel words, phrases, or sentences.

Various problems with the constraint-satisfaction model will be addressed below. The first section takes up the empirical evidence presented in support of the model and argues that it does not address diverging predictions of the constraint-satisfaction approach versus depth-first (one analysis at a time) models of sentence processing such as the garden-path model (Frazier, 1987b; Frazier & Clifton, in press). The second section addresses various methodological problems characteristic of studies conducted in the constraint satisfaction framework. These include (i) the use of temporarily ambiguous sentences always disambiguated to the same structure, which may create an experimental artifact favoring the instantiated structure; (ii) using subjects' plausibility ratings of the experimental sentences or sentence fragments to "predict" processing time for the next word, much on a par with using a human perceiver to classify phonological input, coding the human perceivers judgments into a model, and then claiming the model offers an account of speech perception; (iii) varying the weights in the model to explain different sets of data; (iv) using distinct frequency counts and different correlations on a *post hoc* basis to account for different sets of experimental data, taking significant correlations as support of the model but not taking the nonsignificance of many other correlations as disconfirmation of the model. The third section takes up problems due to the impoverished account of syntactic structure—a problem readily seen in the case of global structure and in the case of languages with head final phrases, but also present, though less readily apparent, in cases of local structure, e.g., when the input is ambiguous with respect to the head of a phrase. Many types of evidence indicate an independence between lexical and syntactic structure. This evidence is discussed in the fourth section. It is problematic for any purely lexical account of syntactic processing—whether the model is developed in the constraint satisfaction framework or in some other framework. The issue of automaticity and language specialization is addressed in the fifth section. It is argued that accounts of automaticity in language processing are difficult to come by if contextual probability, not grammar, is the force driving sentence processing. Though it is argued that the constraint-satisfaction model does not provide a predictive or explanatory model of sentence processing, it might provide a theory of lexical processing. The

sixth section examines this possibility noting that even in the lexical domain the model is only one of several contenders. Further it does not account for the processing of novel words, regularly inflected words, or discontinuous words. The seventh section takes up constraint satisfaction as a model of processing prestored representations, where the model is most promising.

## CONCERNING THE EMPIRICAL EVIDENCE

A large and growing number of detailed empirical studies have been presented which show an influence of specific frequency and probabilistic constraints on comprehension difficulty. With respect to the reduced relative clause (RRC) ambiguity, for example, the relative frequency of the transitive versus intransitive subcategorization of the verb predicted whether subjects had difficulty comprehending reduced relative clauses (MacDonald, 1994). Subjects in her experiment had no difficulty comprehending reduced relatives with *pushed*, *carried*, or other preferred transitive verbs. MacDonald, Pearlmutter, and Seidenberg (1994) pointed out that the relative frequency of past tense versus passive participle categorizations of the verb (V) influenced sentence completion responses in the study of Spivey-Knowlton, Trueswell, and Tanenhaus (1993). Also, the rated semantic plausibility of the reduced relative interpretation influences on-line reading times of temporarily ambiguous sentences disambiguated to their relative clause interpretation (Macdonald, 1994, for example). According to the constraint-satisfaction model, the influence of all these factors is in choosing the appropriate analysis of the lexical items in the first place (how they are put together into a well-formed structure is not entirely clear). The processor's use of these "constraints" is thus attributed to the original phase of processing the sentence, not to revision of an already assigned structure. Moreover, the effect of factors concerning the structurally dispreferred analysis is not limited to the processing of the structurally dispreferred analysis of the sentence, as it is in depth-first theories like the garden-path theory, but should be observed also in the processing of the structurally favored analysis.

With one exception (see discussion of Pearlmutter & MacDonald, 1995, below), the empirical evidence offered in support of the constraint-satisfaction approach has invariably looked at the influence of constraints on comprehension of the structurally disfavored syntactic analysis (i.e., the reduced relative structure and the sentential complement structure, in the studies by Garnsey, MacDonald, Pearlmutter, Tanenhaus, Trueswell, and colleagues). This is unfortunate. In the constraint-satisfaction model, structural and non-structural factors enter into initial analysis of the input. In depth-first models,

nonstructural factors may not guide initial syntactic analysis, but they do guide structure evaluation and reanalysis. Therefore, by examining structures that require reanalysis according to depth-first models, the only clear difference in the predictions of the models depends on being able to tease apart effects of first analysis versus rapid evaluation or reanalysis.

Where the theories make clearly diverging predictions is in the role of certain contextual constraints on the processing of the structurally preferred analysis. For depth-first models, the semantic plausibility of the structurally disfavored analysis should *not* influence comprehension times for sentences disambiguated to the structurally favored analysis whereas in the constraint-satisfaction model it should. For example, the processing of simple main clauses should not be influenced by the semantic plausibility of the reduced relative clause analysis (*The horse raced past the barn*) according to depth-first models such as the garden-path model, because the reduced relative clause analysis is not computed at all. Similarly, the semantic plausibility of the relative clause analysis should not influence comprehension of complement clause sentences (*John told the girl that Bill liked the story*), assuming the complement clause analysis is structurally preferred. However, even according to depth-first models, the semantic plausibility of the simpler (e.g., main clause) analysis may influence the processing of a more complex (e.g., reduced relative clause) sentence, since the parser will first construct and evaluate the main clause analysis before it becomes apparent that the structurally more complex reduced relative analysis is correct. The constraint-satisfaction model does not predict an asymmetry of this sort depending on the relative structural complexity of the various analyses of an ambiguous input. All structures should be syntactically and semantically evaluated in proportion to the evidence favoring the particular analysis. In short, though proponents of constraint-satisfaction models have presented lots of very interesting data showing effects of detailed probabilistic constraints on comprehension times, the studies have not targeted the examples (structurally preferred analyses like the main clause analysis) where clearly distinct predictions are made by depth-first versus constraint-satisfaction (or other breadth first) models.

Arguably Taraban and McClelland (1988) provided a counterexample to the above claim. They showed that PPs which more plausibly modify the NP in a "V-NP-PP" sequence show no reading time penalty even though this is the structurally unpreferred analysis. This is not exactly the circumstance at issue, however. The (im)plausibility of the unpreferred structure would need to influence the reading time on the preferred structure first. Showing that the plausibility of an analysis influences processing times for that analysis is neutral between theories. Moreover, they studied structures which are stored in the lexicon. On anyone's theory, the arguments of a

particular verb and their semantic or thematic roles must be represented. (See discussion of the thematic processor in Rayner, Carlson, & Frazier, 1983; for relevant data also see Clifton, Speer, & Abney, 1991.)

Consider the circumstances where three structural analyses of an input are available, A (the structurally simplest one) and two alternatives B and C, as summarized in (1). It may be significant that garden paths have never been convincingly demonstrated in the processing of analysis A (the structurally simplest one). In principle, this should be possible in the constraint-satisfaction model since lexical frequency and semantic or contextual factors should be able to conspire to favor some analysis other than the structurally simplest one (A). In such cases, if analysis A ultimately proves to be correct, perceivers should show evidence of having been garden pathed by a syntactically more complex analysis even though the syntactically simpler analysis is correct. In other words, a constraint-satisfaction model leads us to expect that all the situations depicted on the right-hand side in (1) obtain.

(1) Garden-path model	Constraint Satisfaction
Analysis A - 1st computed	A = 1st    A = 2    A = 3
Analysis B - 2nd determined by	B = 2nd    B = 1    etc.
Analysis C - characteristics influencing revision	C = 3rd    C = 3

Analysis A = Structurally preferred; 1,2,3 = order in which analyses are computed or their relative activation levels.

One potential demonstration of a garden path in a simple structure is reported in Crain and Steedman (1985) where sentences temporarily ambiguous between a complement clause and a relative clause analysis are placed in biasing contexts. In this study, in an incongruent context, the structurally preferred complement clause analysis led to as much difficulty in a grammaticality judgment task as did the relative clause analysis. But judgments in this experiment may simply have reflected confusion on the part of the subjects, especially since the complement clause sentences began with an expletive *it* which was immediately preceded in the incongruent condition by a noun phrase (NP) of the appropriate syntactic and semantic type to serve as the antecedent of *it*. The effect of this would be to eliminate the expletive analysis of *it*, rendering the complement clause analysis ungrammatical.<sup>3</sup>

<sup>3</sup> In the Crain and Steedman (1985) study, response times were very long and indicated that subjects were not responding until after the entire sentence had been read. This too weakens the claim that the more complex relative clause was computed from the outset in the relative clause context.

As emphasized above, it is completely accidental on the constraint-satisfaction model that garden paths have not been demonstrated in simple syntactic structures (as a result of computing a more complex syntactic structure). On this model, sentences like (2) should garden-path perceivers if all nonstructural factors favor the reduced relative clause analysis.

- (2) a. The children pushed quickly by the armed guards.  
b. The students often taught by the Berlitz method.

It has been claimed by proponents of constraint-satisfaction models that one study does show perceivers are garden pathed in the syntactically simpler structure of a temporary ambiguity, namely in the main verb (MV) analysis of MV-reduced relative clause ambiguities.<sup>4</sup> Therefore, this study will be discussed in detail here.

Pearlmutter and MacDonald (1995) conducted a self-paced reading study using sentences like those in (3) containing a MV/RRC ambiguity and an ambiguous (cooked) or "unambiguous" (bubbled) verb.

- (3) a. The soup cooked in the pot but was not ready to eat.  
b. The soup bubbled in the pot but was not ready to eat.

Forty-eight MIT undergraduate subjects were tested: 16 high-span and 32 low-span subjects, as classified by Daneman and Carpenter's (1980) reading span test. At the preposition (word 4) and the conjunction (word 7) high-span subjects showed longer reading times in ambiguous than in unambiguous sentences; low-span subjects did not show any ambiguity effect. High-span subjects were also more accurate (98.5%) than low-span subjects (93.5%). The experimental results were attributed to differential sensitivity of high versus low-span readers to probabilistic constraints in the stimuli. Off-line rating studies were performed to show that low-span subjects' reading times were correlated with simple lexical frequency constraints, whereas those of high-span subjects were correlated with more complex plausibility constraints.

Specifically, ratings were obtained for the plausibility of the MV intransitive interpretation relative to the MV transitive interpretation using sentence fragments like those in (4). Following Pearlmutter and MacDonald, these will be dubbed the "comparative" rating data. The finding that the

<sup>4</sup> At the CUNY (1994) conference, the Pearlmutter and MacDonald (1995) study was described as a counterexample to my claim that garden paths have not been demonstrated during processing of the syntactically simple analysis of a sentence. In the Pearlmutter and MacDonald paper, they made the more circumspect claim that the study showed an effect of an alternative syntactic structure (presumably the MV transitive) during the processing syntactic structure. However, as we shall see below, there is no clear evidence in their study for even this weaker claim.

comparative rating in (4) correlated with high-span reading times (at word 7) but not with low-span times is the basis for their claim that high-span readers are garden pathed in simple syntactic structures due to use of probabilistic contextual constraints.

- (4) MV transitive: The soup had cooked the  
MV intransitive: The soup had cooked in the

The plausibility of the relative clause (RC) analysis and of the unambiguous control was rated using the fragments in (5) and (6).

- (5) RC: The soup that had cooked in the  
(6) Unambiguous control: The soup bubbled in the

At this point, let us stop to consider the predictions of a constraint satisfaction model. Competition between a syntactically more complex (RRC) and a less complex (MV) analysis should result in long reading times (hence the biggest difference scores between ambiguous and unambiguous sentences) when the complex analysis is most plausible. This is because frequency will favor the simple MV structure and plausibility will favor the complex RRC structure. Yet no effect of this sort is reported even for the high-span subjects who are supposed to be particularly responsive to contextual probabilities. Instead stepwise regression analysis indicated that at the preposition (word 4) high plausibility RRC sentences yielded a *smaller* ambiguous-unambiguous difference than the low-plausibility RRC sentences did. In other words, the data disconfirmed the predictions of a competition model. Pearlmutter and MacDonald (1995) commented only that the presence of the preposition at this position is consistent with a reduced relative interpretation and inconsistent with the MV *transitive* interpretation of the ambiguity; thus, high-span subjects' reading times were shorter when a constraining word (the preposition) promoted a plausible interpretation. (Underscoring is mine.)

What is going on here? The disconfirmation of the hypothesis that sentences with a plausible RC rating would lead to long processing times for the high-span subjects due to the competition between two grammatical analyses (MV intransitive and RRC) is being ignored. Instead the RRC is being compared to the now ungrammatical MV *transitive* analysis to explain the evidence incompatible with a competition framework. In short, the results do not confirm the hypothesis that the RRC analysis interferes with the processing of simple main clause sentences.

The aspect of their data that impressed Pearlmutter and MacDonald (1995) was that high-span subjects did not show simple lexical frequency effects but effects of combinatorial constraints of a particular noun & verb combination. At word 7 (*but*), high-span subjects showed a larger difference

between ambiguous and unambiguous sentences for those sentences with a more plausible transitive structure and a smaller difference for those items with a more plausible correct intransitive structure. The authors admitted that this result did not show whether the high span subjects are truly using plausibility information or whether they are instead sensitive to lexical frequency information that is less available to low span subjects. For example, the relative plausibility of the MV intransitive and the MV transitive interpretation is likely to be correlated with the relative frequency of MV intransitive and the MV transitive uses of the ambiguous verbs in English (Pearlmutter & MacDonald, 1995). An additional rating study (Experiment 2) was conducted to investigate this issue.

Experiment 2 correlated Experiment 1 reading time data with the rating data described above (supplemented with rating data from University of Illinois subjects) and with the frequency of five different argument structures. The results showed that the difference score at word 7 correlated better with the comparative rating [see (4)] than with a corpus based frequency count of the cause-theme argument structure (e.g., *The soup cooked the vegetables*.) The difference score at word 7 did not correlate at all with the frequency of the other argument structures tested: agent only (*John cooked*), theme only (*The soup cooked*), agent-theme (*John cooked the vegetables*), or passive (*The soup was cooked by John*). The correlation between reading times at word 7 for high spans and the comparative rating shows, Pearlmutter and MacDonald (1995) thought, that combinatorial constraints must be exploited by high-span readers. The absence of strong correlations with argument structure frequency showed that simple lexical frequency was not responsible.

Let us review the evidence that perceivers are garden pathed in syntactically preferred structures.

- (i) Did the syntactically complex RRC structure interfere with the processing of simple MV sentences? No. Counter to the predictions of the competition model, less difficulty was evidenced in sentences with plausible RRC structures than in sentences with less plausible RRC structures.
- (ii) Did the syntactically simpler structure take longer to process when it was temporarily ambiguous than when it was unambiguous? Not for two-thirds of the subjects.
- (iii) For the high-span subjects who showed an ambiguity effect at word 7, was the size of the effect best predicted by the frequency of the argument structure of the verb (= word 3)? No, the comparative rating did a better job.

Though (i) and (ii) argue against a constraint-satisfaction model employing competition between alternative syntactic analyses of an input, perhaps (iii) provides some evidence in support of the model. To determine whether it does, one would ideally like some predictions in advance of the data.

How can the garden-path model account for (i) through (iii)? The absence of a garden path in simple main clause sentences for most readers [see (i) and (ii)] follows immediately from the prediction that the simple main clause analysis is initially pursued by perceivers. But what about (iii)? Indeed, why for that matter should high-span readers show an ambiguity effect at position 7 at all?

Imagine that high-span readers are simply more motivated than low-span readers in both the memory span task and the comprehension task. This would explain why there existed a correlation between performance on the "memory" task and the comprehension task. It also explains why high spans showed higher accuracy on the comprehension questions than low spans did. Notice that the accuracy difference between high- and low-span readers was mysterious on the constraint-satisfaction account, which predicted high spans would consider several structures whereas low spans would consider only the correct one. Why should higher accuracy not be found for the low spans who considered only the correct structure?

To account for (iii), one might assume that more motivated readers are more likely to show plausibility effects at structurally determined points in the sentence. Notice that the comparative rating was the most relevant plausibility rating including the verb and preposition from the ambiguous sentences. It is not surprising that the plausibility of the relation between a verb and a following prepositional phrase (PP) should be available and still show up at the first word (*but*) securely marking the end of the first clause in the experimental sentences.<sup>5</sup> On many accounts, interpretation effects are expected at clause-final position. If high-span readers are more motivated readers, as suggested by the accuracy data, then it is not unexpected that their processing times will be linked more closely to the detailed structure of the input than the processing times of less-motivated low-span subjects. In short,

<sup>5</sup> As Chuck Clifton pointed out to me, the plausibility of the intransitive main clause (MC) reading did not vary much, suggesting that variation in the plausibility of the transitive MC reading may have been responsible for the variation in the comparative rating. In the absence of separate transitive MC versus intransitive MC ratings, this is difficult to assess. Further, the fact that the correlations were computed using reading times from only ten sentences raises the possibility that just one or two items might have been responsible for the observed correlations. This introduces further uncertainty about the correct interpretation of the word 7 correlations for the high-span readers.

there exists no strong evidence to date of perceivers' having been garden pathed by a syntactically more complex structure during the processing of the syntactically preferred (simple) analysis of a sentence.

### METHODOLOGICAL CONCERNS

Recent experimental studies of sentence processing, including those working in the constraint-satisfaction framework but also others (e.g., Ni & Crain, 1990), have investigated sentence processing using temporarily ambiguous sentences always disambiguated to one and the same syntactic structure. This is poor experimental design. Experimental subjects are likely to catch on to the fact that, for example, every sentence beginning with *only* continues with a (reduced) relative clause structure. Even without overt markers like *only*, repeatedly disambiguating a temporarily ambiguous sentence type toward one particular structure may influence subjects' behavior.<sup>6</sup>

For many years it has been apparent that the processing time for a word, phrase, or sentence is influenced by both syntactic structure and semantic/pragmatic plausibility of a sentence fragment (e.g., Forster & Olbrei, 1973, to take one example). The question is how this comes about. Is it due, say, to syntactically integrating a word into context and then interpreting the resulting structure or is it due to a constraint-satisfaction process involving structures competing as functions of their current activation states? Either mechanism could explain why a word is processed faster in a context with strong biases.

Constraint-satisfaction model proponents often take demonstrations of frequency and contextual probability effects to favor their model. For example, subjects will be asked to rate the plausibility of sentence fragments continuing with a word disambiguating the fragment toward one or another syntactic structure. If reading time in an experiment then varies as a function

<sup>6</sup> As long as I am griping about recent experimental practices, I would also like to fuss about presenting data only in terms of difference scores, e.g., between ambiguous and unambiguous sentences. Difference scores can be illuminating by highlighting theoretically crucial aspects of the data. But a full presentation of the original data is also needed in order to ensure that differences are due to variations in the experimental condition and not the baseline. For example, comparing reduced relatives to unreduced relatives, large difference scores might emerge because of the predicted complexity of the reduced relatives (*The horse raced past the barn*) or due to an exaggerated estimate in the case of unreduced relatives (*The horse that was raced past the barn*) due to processing operations completed during the processing of the disambiguation (*that was*). Determining the source of the difference scores is theoretically important and can be independently assessed only if absolute processing times are presented in addition to the difference scores based on the actual reading times.

of the plausibility ratings, this is viewed as evidence supporting the constraint-satisfaction model. After all, the constraint-satisfaction model explicitly claims that frequency and contextual probability should influence sentence processing. Indeed, but how? For the constraint-satisfaction model proponents to take plausibility ratings generated by human perceivers for particular sentences and use this as the basis for the model's predictions about processing those *same* sentences is unimpressive and begs the very question they set out to answer. A true model predicts what humans will do and explains how they will do it using the model itself (not a human perceiver) to generate its predictions.

Further, for any model to account for just one set of experimental data and then use different ratings and different sources of constraint to account for a new set of data is not impressive. For example in the Pearlmutter and MacDonald (1995) study, why was the absence of a negative correlation between the frequency of the passive participle use of a verb and reading time for a main clause sentence containing that verb not disconfirmation of the hypothesis that the RRC structure interferes with processing a main clause sentence? In other studies, passive participle use of the verb has correlated with reading time of the RRC analysis (see MacDonald, Pearlmutter, & Seidenberg, 1994). So the prediction would seem to be that high passive participle frequency facilitates processing of an RRC sentence *and* it interferes with processing of a main clause (MC) sentence. If the second prediction is not confirmed, why should this not count as disconfirmation of the model? The goal, shared I believe by everyone engaged in the psycholinguistic study of human sentence (or language) processing, is not a theory of the data generated by one experiment.<sup>7</sup>

How can rating data generated by humans be used in a principled way to set the parameters or motivate the predictions of a constraint-satisfaction model? One crucial step is to let ratings set parameters of the model and then test the model on a *different* set of sentences. Another crucial step is to predict in advance (*not post hoc*) which correlations between ratings and experimental data should be significant. Those predictions should then be tested on the results of more than a single experiment.

According to the view held by those working in a constraint-satisfaction framework, all the action in sentence processing—the heart of the problem—concerns the processor's use of frequency and contextual probability information. Therefore, it would seem incumbent on those pursuing this approach to explicitly specify in advance of an experimental test of the theory which frequencies should matter and why. This matter is two-sided.

<sup>7</sup> Indeed, in my opinion, the goal should not be to model the data at all, but to model the representations and processes that underlie our ability to comprehend sentences.

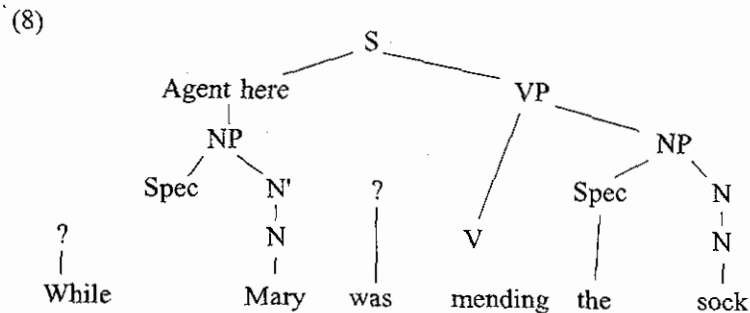
On the one hand, proponents of constraint-satisfaction models, and others (see Gibson, Schutze, & Salomon, 1995; Merlo, 1994; Mitchell, Cuetos, & Corley, 1995), deserve credit for empirical attempts to sort out which frequencies matter. On the other hand, the resulting evidence, though not predicted in advance, has often been taken as evidence favoring a constraint-satisfaction model. Surely such claims are premature in the absence of an articulated account of which frequencies should matter. If it is contingent frequency, then the model should explicitly define contingent frequency. Is it computed over a two-word window, a five-word window, a 25-word window? Is it contingent frequency given a string of syntactic categories, or does the higher syntactic structure matter? Or is it computed over semantic categories and, if so, which ones?

### THE TREATMENT OF SYNTACTIC STRUCTURE

Much evidence has been taken by proponents of depth-first models to support the existence of general structure-based parsing strategies like *minimal attachment* (see Frazier, 1987b, for example). This evidence is explained, MacDonald, Pearlmutter, and Seidenberg (1994) have claimed, by the frequency of the syntactically simpler structure. But, given their discussion of processing, the only structure assigned is prestored lexical structure, which they assumed includes lexical category and the immediate X' projection of lexical items. Take, for example, the late closure ambiguity in (7) which has been discussed extensively in the literature.

(7) While Mary was mending the sock fell off her lap.

The syntactic structure which their theory assigns to the first clause of (7) seems to include a lexical category label for each word and the syntactic structure in (8).



There is no representation in this structure of the subordinate clause object versus main clause subject ambiguity nor a representation of the frequency

of the two analyses because there is *no* relation at all between the subordinate clause and the main clause.

Several points must be emphasized. First, there is no representation of global syntactic structure in this theory. Second, the representation of the structure that goes beyond the immediate X' projection of lexical heads is nontrivially deficient. For example, the theory *only* permits clauses to have a subject by listing for each verb separately each of the thematic roles of the subjects it may occur with. This is a problem because some subjects do not receive a thematic role. (*It's raining, It seems John left, There's no fresh water*). Further, many generalizations apply to all subjects regardless of their thematic status, such as tag-question formation (*There's a problem, isn't there?*) In short, syntactic structures are available only to the extent that they are already listed. And global structure beyond the X' projection of lexical items is not represented at all.

Proponents of constraint-satisfaction models suggest that the existence of general syntactic preferences (for simple structures) may really be due to the high frequency of the preferred structures, e.g., the subordinate clause object versus main clause subject analysis of an NP following a verb in an initial subordinate clause. But this is problematic. If the structures are not represented, how can their frequencies matter? Let us assume that only the overall frequency of a verb's transitive versus intransitive usage is represented. (If "combinatorial" information is also listed, we must know what combinatorial information is listed and what is not listed. In the absence of such an account, the model makes no predictions.) If no relation between the subordinate clause and the main clause is computed, then presumably the relative frequency of a postverbal NP being object within a subordinate clause versus subject of a main clause is not represented and, consequently, not available as an explanation for the structural bias leftover once lexical biases are eliminated. Further, a frequency account does not take seriously the cross-language finding: The *same* principles seem to operate in processing different languages. If the frequencies of the distinct analyses of temporarily ambiguous sentences happen to universally favor the same structures, this in itself calls for an explanation. On the frequency account, the facts are completely surprising. Why should the frequency of past participles in one language correlate with the frequency of past participles in another language where the syntactic distribution of past participles may be more restricted in terms of the number of constructions in which the past participles participate? Moreover, in some languages (e.g., Dutch) simple past tense forms are rarely used by speakers, who are much more likely to use perfect forms of verbs. Clearly these sorts of factors—the distributional freedom of a form and the systematic preference for one structure over another by speakers—differ across languages. Though the expressive needs

of speakers may or may not remain constant across languages, any constancy that may exist does not by itself entail commonality across languages in the structures used to convey a particular message, especially commonality as encoded by the relative frequency of stored lexical forms.

The constraint-satisfaction model also fails to offer an account for certain well-established syntactic processing generalizations. For example, the active filler strategy (or the more general minimal chain principle of de-Vincenzi, 1991, 1995) specifies that displaced constituents are assigned to the first available position consistent with the grammar of the language being analyzed (*not* the first available NP slot given local constraints). The active filler strategy has been argued to apply in English (Frazier & Clifton, 1989), Dutch (Frazier, 1987a; Frazier & Flores d'Arcais, 1989), German (Schriefers, Friederici, & Kühn, 1995), Spanish (Gilboy, personal communication), Japanese (Nagai, 1995), and Korean (Koh, personal communication). The principle cannot even be stated in terms of the syntactic processing theory offered by the constraint-satisfaction model (e.g., the one by MacDonald et al., 1994, discussed in the paragraph above). The difficulty is that higher-level structure is needed to identify many displaced constituents (fillers), and to identify appropriate gaps (traces) for them. See discussion of example (9) below for an indication of why the problem extends far beyond the failure to capture the active filler strategy preferences *per se*.

In addition to the problems resulting from the absence of global syntactic structure, many problems result from using lexical items to project structure. The problems fall into two classes. One class concerns the circumstances where lexically unpreferred hypotheses are favored. For example, Frazier and Clifton (1989) showed that the processor abides by the active filler strategy *even* when this means assigning a filler to a gap corresponding to a lexically unpreferred position, i.e., to the less frequent subcategorization of a verb. See Speer (1995) for late closure effects with frequency balanced verbs; see Gibson et al. (1995), Traxler and Pickering (1995), and Adams, Clifton, and Mitchell (1995), for additional examples where the preferred syntactic analysis does not coincide with the lexically preferred analysis. It often appears that lexical frequency determines which initial analysis of a sentence is pursued. For evidence that this appearance results from the slowness of button-pressing techniques, see Kennison (1995). For evidence that combinatorial information does not suffice to determine syntactic analysis, see Clifton, Kennison, Albrecht, and Frazier (1995).

The second class of problems resulting from projecting structure lexically concerns incorrect predictions about the consequent delays of analysis in head-final phrases. Frazier (1987a) argued that the Dutch VP node is available before the (phrase-final) verb is encountered. Bader and Lasser

(1994) showed that the German VP node is available before its head is encountered. Moreover, they also showed that an NP is not necessarily preferentially analyzed as being licensed by, or an argument of, the first following verb. Under the circumstances they tested, the preferred attachment site for the NP was the VP headed by the higher (sequentially later) verb. A lexical approach to syntax incorrectly predicts the existence of a preference for attachment to whatever structure is headed by the earlier of two subsequent heads of phrases. In short, the idea of projecting structure bottom up from lexical items meets with several serious empirical difficulties: It cannot explain syntactic preferences that do not coincide with lexical preferences, it incorrectly predicts delays of analysis in head-final phrases, and it leads to the expectation that local structure (local licensing of an argument) will always be preferred since no nonlocal structure will be available to provide the alternative analysis or attachment site.

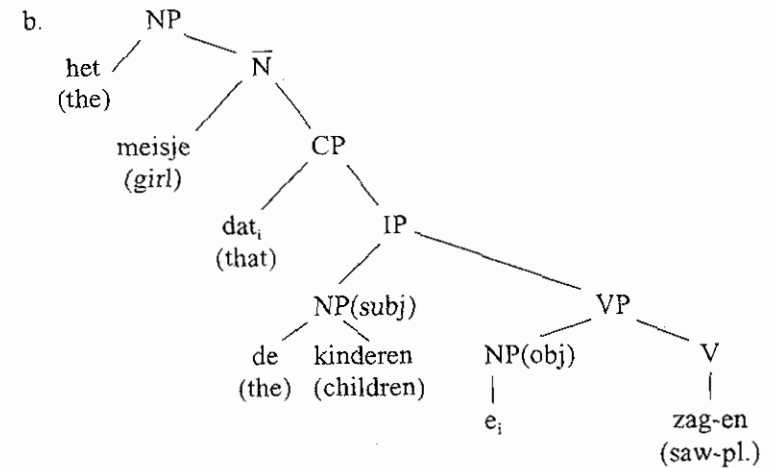
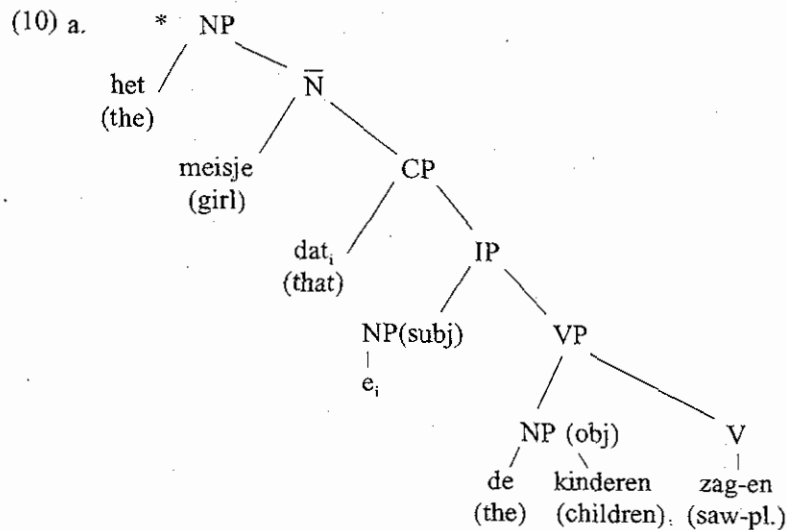
It was emphasized above that the predicted delayed analysis of head-final phrases has been disconfirmed empirically. Clausal structure is also predicted to be postulated late, only at the verb. This has ramifications beyond those immediately concerning the analysis of the head final phrase itself. The absence of clausal structure early in processing necessitates late identification and assignment of preposed constituents. This is a problem for languages like English, where priming is observed at the gap from which a phrase has been moved even if the verb of that clause has not yet been encountered (see Zurif, Swinney, Prather, Solomon, & Bushell, 1993, for a study of subject relative clauses in English.) Available data from V-final languages also strongly suggest that preposed phrases are assigned to gap positions long before clausal structure can be projected from the verb. For example, the active filler strategy together with minimal revisions [discussed in more detail below, see (11)] predicts that (9a) is the preferred structure of the Japanese fragment in (9).

- (9) a. Hanako<sub>i</sub>-wa [e<sub>i</sub> [NP-ga  
           |          |                  |  
       Proper Name Topic              Nom  
       b. Hanako<sub>i</sub>-wa [NP-ga[e<sub>i</sub>  
                       |          |                  |  
                   Proper Name Topic              Nom

This is because the filler is assigned immediately to the highest subject position as in (9a), even before NP-ga is processed. The resulting structure (9a) is maintained (by minimal revisions) once NP-ga is processed and therefore NP-ga is analyzed as the subject of a new clause. But, in the constraint-satisfaction model, no structure will be available yet since the subject

position where the gap ("e<sub>i</sub>") sits in (9a) is not postulated until much later. Recall that the subject slot is prestored in a clause that is made available by the verb (crucially, since only a thematic role compatible with the particular verb can appear as subject, and in this theory it is only through this thematic role that the subject argument is "bound" to this slot). See Nagai (1995) for evidence that (9a) is preferred. Comparable examples have been studied in various Dutch and German structures, namely, cases where syntactic ambiguities are resolved earlier in reality than is predicted by this model (see below for relative clause examples). The problem in these cases is not just that the constraint-satisfaction model incorrectly predicts the existence of a delay in the syntactic analysis of the input. The problem is that there is no explanation for the preferred analysis if decisions take place at the late point required in models which project structure from lexical heads. In (9), the processor would already have encountered a nominative-marked NP to serve as subject of the clause by the time it made a decision about the analysis of the initial NP. Thus the topic (*wa*) marked NP should be taken to be the accusative object of the highest clause.

Consider Dutch and German relative clauses disambiguated by number information carried by the clause-final relative clause verb. On-line application of the active filler strategy results in a gap in subject position as shown in (10a). Due to subject-verb number agreement, this analysis will become ungrammatical when the plural verb is encountered—requiring the processor to construct the grammatical (10b), where both the subject and the verb are plural.



The processor would receive disambiguating number information on the clause-final verb before any syntactic analysis was predicted to be assigned on the constraint-satisfaction view. Recall that the sentence node is introduced only by the verb in the constraint-satisfaction model. Hence it is a complete surprise in terms of the constraint-satisfaction model why a subject preference emerges in disambiguated sentences and why disambiguating information on the verb would be ignored by the processor, as it sometimes is (see Frazier, 1987a). In models where top-down syntactic information is used to assign syntactic structure immediately, the presence of misanalyses due to early structuring of the input is expected. Overlooking information incompatible with the assigned syntactic structure is also not mysterious in terms of this view. Of course constraint-satisfaction proponents might give up the claim that syntactic processing proceeds by lexical activation of structure stored with the head of a phrase. But, in that case, they surely owe us some account of how sentence processing does take place.

Let us turn now to the constraint-satisfaction model predictions concerning local structure and how the head of a phrase is identified. Recall that lexical items are stored with their X' projections. Therefore, each noun (N), for example, comes supplied with a dominating N' and N<sup>max</sup>(NP). Given two nouns, each will have an X' structure provided by the lexicon. How does the processor determine whether two distinct noun phrases are present, or two conjuncts of a conjoined phrase or simply a novel compound noun? In the latter case, how is the X' projection of the first noun deactivated? Without a principle granting preference to simple structure, by what means

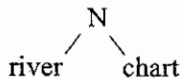
is the lexically provided NP dominating the first noun suppressed? If syntactic structure is supplied by heads of phrases, some mechanism is needed to identify the head of a phrase. At present, the constraint-satisfaction model seems to incorrectly predict a preference for *maximal* structure in novel N-N-N sequences, because each noun is already listed with its X' structure.

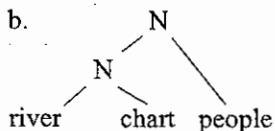
The data concerning processing of novel N-N-N sequences are discussed below. Before turning to them, let us contrast the predictions of the garden-path model with the predictions of the constraint-satisfaction model. In the garden-path model, initial commitments are made to the currently simplest syntactic structure followed by the minimal revision required to rectify any error detected and maintain compatibility with the available information about the input. In a competition framework, where all analyses are considered from the outset, it is unclear why the ultimately chosen analysis of a sentence should bear a *systematic* relation to the syntactically simplest analysis, namely, the relation described by minimal revisions (see Frazier, 1990), given in (11).

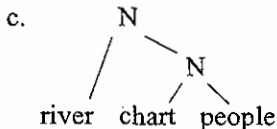
(11) Minimal revisions: Do not make an unnecessary revision.

When revision is necessary, make the minimal revision, maintaining as much of the already assigned structure and interpretation as possible.

The difference between models may be illustrated concretely with respect to the processing of *novel* English N-N-N compounds. A left-branching preference (12b) results from making the minimal revision to the existing N-N analysis (12a); the alternative (12c) involves revision of the existing structure and interpretation (Frazier, 1990). But if all grammatically available structures compete, then with ambiguities such as (12) it is unclear why the preferred analysis should be the one resulting from minimal revisions.

(12) a. 

b. 

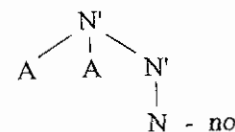
c. 

Similar problems arise for a constraint satisfaction account of Inoue and Fodor's (1995) example (13) involving the preferred interpretation of Japanese prenominal adjective sequences:

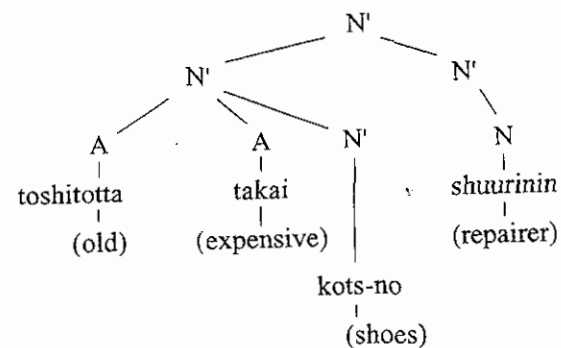
(13) toshitotta takai kots-no shuurinin  
'old (animate) expensive shoe-GEN repairer'

- a. repairer of old expensive shoes  
b. old repairer of expensive shoes  
c. old expensive repairer of shoes

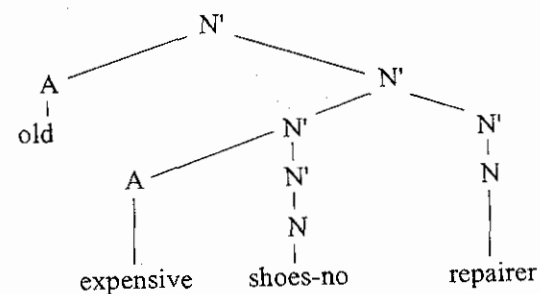
(14) i.



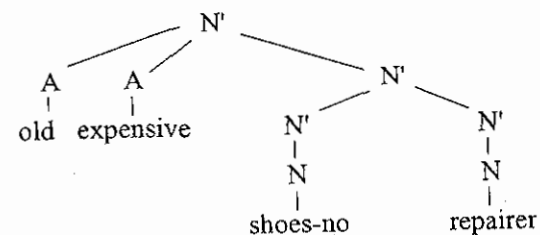
ii.



iii.



iv.



Assuming that syntactic analysis proceeds on-line roughly as each word is encountered, the initial sequence of (13) will be structured as shown in (14i) corresponding to the paraphrase in (13a). When *shuurinin* is encountered, the minimal revision of (14i) is that shown in (14ii) where the structure and interpretation assigned in (14i) is preserved, as shown by the circled con-

stituent in (14ii). According to Inoue and Fodor, this is indeed the preferred analysis. The alternative, where one adjective modifies *shoes* and one modifies *repairer* as in (13c), illustrated in (14iii), might be disfavored by the parser simply by minimal attachment since the analysis entails an extra syntactic node (N') not needed for the analysis in (14ii). (Here I am following the structure provided by Inoue and Fodor for these NPs.) The alternative where *both* adjectives modify *repairer* does not require extra nodes to be postulated, as may be seen in (14iv). This analysis only requires more revisions than the one shown in (14ii) and thus it is correctly predicted to be disfavored by the minimal revisions principle.

If all analyses were considered in parallel, as predicted by the constraint-satisfaction approach, the preference remains mysterious. Minimal revision of the simplest syntactic analysis is a theoretical notion that is apparently not easy to reconstruct in that framework. Of course, frequency might favor the chosen analysis—but, if so, why? Why should the frequency of Japanese prenominal modifier sequences and the frequency of English left-branching versus right-branching N-N-N compounds be related, if indeed frequency factors are responsible for these preferences? Of course, for proponents of the constraint-satisfaction model, the apparent generalization is only an accident or an epiphenomenon. The problem, however, is that taking this approach, everything is an accident. The investigator is no longer encouraged to seek new generalizations or to explain already identified ones. In short, there are few if any problems, questions, or generalizations. There are only frequencies, and we are not told which ones matter or why it is particular frequencies that are important and not others.

A few further concerns will be noted about the current inadequacies of the constraint-satisfaction model as a theory of sentence processing. The model does not specify how the processor arrives at a syntactic representation, or how ungrammatical analyses are excluded. It is unclear how adjuncts will be processed. Adjunct ambiguities are fully nonlexical in the sense that they do not interact with the description assigned to individual lexical items. Thus the constraint-satisfaction model fails to handle most instances of PP attachment, adverb attachment, conjunction, relative clause attachment and adjectival phrase attachment, to take just a few examples.

### INDEPENDENCE OF LEXICAL AND SYNTACTIC INFORMATION

Another class of problems for theories relying on lexical projection of syntactic structure concerns the independence of lexical and syntactic information in many psychologically real phenomena. These include the assign-

ment of syntactic structure to phrases headed by unfamiliar or novel words, or to strings of nonsense words (Epstein, 1961), or to semantically anomalous prose where lexical requirements are violated (Marslen-Wilson & Tyler, 1980). Pure syntactic priming effects in production are also problematic, such as those lexically insensitive effects demonstrated by Bock (1989). In each case, general syntactic configuration either conflicts with lexical specifications or, at least, must be independent of the properties of the lexical items or nonsense words contained in the construction.

Code switching provides a particular powerful case in point. It is a naturally occurring phenomenon in which syntactic structure follows the matrix-language regularities even when the content item heading a phrase is supplied by the embedded language which may have a distinct word order. For example, Joshi (1985) presented examples like (14) which showed a Marathi (V-final) verb phrase (VP) headed by an English verb. The English verb *paint* does not assign case or a theme role to a preceding NP. Thus the English verb cannot be responsible for projecting the VP dominating the direct object and the verb in (15).

- (15) mula khurcaya paint kartat (Marathi VP, English V)  
 boys chairs do + tense

These sorts of examples suggest that syntactic structures have a reality of their own, apart from the words inserted in the structures.

A variety of recent ERP studies have also suggested the existence of a stage of syntactic processing independent from semantic predictability or anomaly. Osterhout (1994 and references therein) argued that the P600 reflects syntactic violations but not other sorts of violations. For example, Osterhout and Holcomb (1992) showed that a P600 component follows the presentation of the word *to* in (16b) compared to (16a) where no syntactic anomaly occurs. McKinnon and Osterhout (1994) showed the presence of a similar P600 in sentences containing subadjacency (17) or ECP (18) violations.

- (16) a. The broker hoped to sell the stock.  
 b. \*The broker persuaded to sell the stock.  
 (17) \*I wonder *which of his staff member*, the candidate was annoyed  
 \*[when his son was questioned by \_\_\_\_\_].  
 (18) \*John, seems that it is likely \_\_\_\_\_, to win.

The P600 attends syntactic difficulty due to a syntactic violation or a need to reanalyze the syntax (see Brown, Hagoort, & Vonk, 1995; Hagoort & Brown, 1994). The P600 is apparently present even in sentences containing semantic biases toward the correct syntactic analysis (Friederici, Mecklinger, Steinhauer, & Hahne, 1995). Difficulty semantically integrating a

word into its context also reveals ERP effects but they show up as a negative shift, the so-called "N400." N400s are large for anomalous words and smaller for more predictable words with a better semantic fit into the sentence fragment. It is unclear how a constraint-satisfaction approach would account for the apparent dissociation of syntactic effects as indicated by the P600, and lexical-semantic-contextual predictability effects, reflected in the N400. Such dissociation during processing should presumably be impossible on that theory.

### AUTOMATICITY AND LANGUAGE SPECIALIZATION

The appeal of constraint-satisfaction models seems in large part to stem from the use of a single (parallel distributed) processing architecture for all human cognitive processing (perhaps for all processing period). It was suggested in MacDonald et al. (1994) that, while linguistic representations may be distinct from nonlinguistic ones, the processing of those representations and the ambiguity resolution principles may not be particular to language at all. This seems plausible and may well be correct in some sense. But to take this speculation seriously it is necessary to explain the automaticity of language processing and the considerable evidence suggesting human specialization for language and language processing. It may well be that in *all* areas where humans are specialized for computing certain kinds of representations (e.g., visual, motor, etc.) specialized architectures reflecting the structure of the representations are found. In this sense, having a specialized architecture for the language processing mechanism would reflect the general case, and be nothing special at all. Certainly in other domains where humans are specialized for computing particular kinds of representations, specialized systems are the norm. Consider the mere existence of visual cortex or visual association areas, for example. They hardly reflect unspecialized systems. Indeed, as Geschwind concluded, there really is no shred of evidence for the existence of unspecialized systems, or general-purpose computational systems, in the human brain. Consequently, the generality of the processing architecture of constraint-satisfaction models, while superficially attractive, may actually constitute an argument against the model.

### CONSTRAINT SATISFACTION AS A THEORY OF LEXICAL PROCESSING

It has been argued repeatedly (most eloquently by MacDonald et al., 1994) that syntactic processing and syntactic ambiguity resolution is accom-

plished by the same mechanisms as lexical processing. It is thus relevant to assess the constraint-satisfaction model as a model of lexical processing. At the outset, however, it should be stressed that the inadequacies of the model as a theory of lexical processing have a somewhat different status than its inadequacies as a model of sentence processing because in some cases no better or alternative lexical processing model is available. This stands in sharp contrast to the situation with syntactic processing where far more developed, detailed, adequate, and explanatory models exist. Indeed, most of the questions posed in terms of other models of sentence processing cannot even be raised in the constraint-satisfaction model because it is presently so impoverished, consisting primarily of the claim that frequency and context matter. In any case, it seems relevant to evaluate constraint satisfaction in terms of its strengths as well as its weaknesses, including as a model of lexical processing.

As a theory of lexical processing, the constraint-satisfaction model exhibits several drawbacks:

- (i) Because the model fails to say how parts are put together into larger constituents, it fails to deal with discontinuous dependencies such as complex verbs in Dutch where the verb and its prefix (which may have a noncompositional and unpredictable meaning) may be separated by theoretically unbounded material (see Frazier, Flores d'Arcais, & Coolen, 1993). Further, as in *all* models that use activation levels of a single set of lexical units to identify a word W1 and a following word W2, the model does not explain why a word intervening between the two parts of a discontinuous word does not eliminate the activation levels that would help identify the second part of the discontinuous word.
- (ii) Arguments exist showing lexical and syntactic systems must be separate since they may entertain *contradictory* analyses of the same input (Frazier et al., 1993).
- (iii) Regularly inflected words (and morphologically complex words derived from productive level 2 affixes) are apparently treated as prestored in the constraint-satisfaction model, despite evidence of active analysis by perceivers of English (see Carter, 1994; Marcus et al., 1993; Opler, Robinson, Kaufman, & Satake, 1994, for examples). Also it is implausible to treat all words as prestored: Consider languages where a word has been estimated to exhibit over a million distinct forms, as in the example of a Turkish verb (see Hankamer, 1989).

- (iv) Novel words, ones never before encountered, are left untreated.<sup>8</sup>
- (v) Bottom-up priority is proposed as an explanation of apparent effects of modularity. The association between the form and the meaning of a word is stronger than the association between context and meaning. This—and this alone—is supposed to account for apparent effects of modularity in lexical processing. But this account is contradicted by Seidenberg, Tanenhaus, Leiman, and Bienkowski's (1982) evidence showing multiple access of all lexical analyses (N, V) in a syntactically unambiguous context. Here multiple analysis was observed despite a 100% contextual constraint (*\*They all [rose]<sub>Noun</sub>*). Compare this to selective access for noun-noun ambiguities in *weaker* contexts. According to bottom-up priority, only the strength of the contextual constraint should have mattered.

Assuming that a theory of lexical processing must account for how novel words are recognized and must offer an explanation for the properties of the lexical processing system (why some forms are stored, and others not, why lexical processes are basically bottom-up but apparently do not reflect merely the degree of association of form and meaning, etc.), the constraint-satisfaction model fails as a model of word processing for many

<sup>8</sup> An anonymous reviewer suggested that, under a "constructional account", syntactic structures are themselves paired with meaning. "This would enable (1) syntactic structures with meaning to have frequency, and (2) for novel words to be 'recognized' by virtue of the fact that they occur in a construction which itself involves meaning and is also involved with a limited set of semantic classes of verbs." This is an interesting suggestion but it does not truly address the problem of how predictable aspects of the structure or meaning of novel words are computed. For example, native speakers of English know that you can "rethink the problem" but not "rethink about the problem." Likewise for all verbs derived with *re-*: a simple direct object is grammatical despite the fact that the base verb cannot take one (*think the problem*) and oblique arguments are not possible even if the base verb allows them (*think about the problem*). This information must be computed for novel inputs, like *reznok the villain*, it is not available from a lexical entry for *reznok*.

To take another example, a novel compound word like *womantrader* may mean either a female trader or someone who trades woman, even in a null context. It may not mean someone who trades with women (in a null context). This is because *woman* may be assigned the theta role theme but not goal. Therefore, the only way to get the meaning *someone who trades with women* is to derive it, not grammatically, but from a salient property of context. In short, predictable aspects of the structure and meaning of novel words must be computed. In this sense, recognizing novel words is like parsing sentences: The parts must be put together in grammatically constrained ways. A theory which does not tell us how parts are put together consequently does not adequately treat the processing of novel words, whether it appeals to construction grammar or some other type of grammar.

of the same reasons that it fails as models of syntactic processing: The model essentially does not treat novel items or discontinuous dependencies but only makes explicit claims about the activation of prestored representations. I now turn to the issue of whether constraint satisfaction provides a good model of processing prestored representations.

### THE ADEQUACY OF THE CONSTRAINT SATISFACTION AS A MODEL OF PROCESSING PRESTORED REPRESENTATIONS

As emphasized above, the constraint-satisfaction model really offers only a theory of how prestored representations are processed. Is it a good model of processing prestored representations? For example, does it offer the correct account of ambiguity resolution once its domain is limited appropriately to the area (prestored representations) where it is most plausible and where it makes explicit claims?

MacDonald et al. (1994) and Trueswell and Tanenhaus (1994) argued that lexical ambiguity resolution for prestored elements is frequency sensitive and shows context effects and in these respects it is claimed to be like syntactic ambiguity resolution. These claims are intended to challenge the notion that prestored representations (many lexical representations) are accessed in parallel whereas computed representations such as syntactic representations are computed essentially one at a time.

Whether the existence of frequency and context effects in both lexical and syntactic ambiguity resolution mechanisms suggests the operation of similar processing mechanisms in the two cases depends on the correct account of these effects.

MacDonald et al. (1994) argued their case by pointing out that lexical access is not fully parallel: Access depends on the relative frequency of the stored alternatives and, sometimes, on context. Well-known cross-modal priming studies argue for a system with multiple access of alternatives (Seidenberg et al., 1982; Swinney, 1979). Evidence from eye movement studies converges on this conclusion. However, as noted by MacDonald et al., this conclusion is limited in two respects. The relative frequency of the stored alternatives matters. Strongly biased words do not show truly simultaneous access of the dominant and subordinate sense (Simpson, 1985, Rayner & Duffy, 1986, 1987; Simpson & Burgess, 1985). Also, certain contexts favoring a dominant sense have been claimed to eliminate activation of the subordinate sense (Tabossi, Colombo, Job, 1987, Duffy, Morris, Rayner, 1988; Tabossi, Colombo, & Job, 1987; but also see Love & Swinney, 1995).

Two descriptions of these facts are available. The "reordered access" model allows contexts to direct lexical access, eliminating difficulties due

to ambiguity (seen, for example, in the long fixation times on balanced ambiguous items in neutral contexts). This mechanism is completely consistent with the constraint-satisfaction model, since frequency and context can interact to identify a single candidate.

The alternative description of the effects is the selection model (Rayner & Frazier, 1989). In this model, lexical access procedures are not directed by extralexical context, only later selection mechanisms are. Context may speed activation by priming features of a lexical representation (Tabossi et al., 1987) or by allowing rapid integration with the sentence context. The effects of frequency in biased items are due to the processor using the activation or access time differences themselves as evidence about the appropriate analysis: If one analysis becomes available much sooner than an alternative, it may be accepted, or even integrated with context, before the alternative becomes available. Presumably in all models, acceptance of one analysis inhibits alternatives. This selection account predicts long fixation times on balanced ambiguous items in neutral contexts since the processor has little evidence to choose between alternatives: Uncertainty is reflected in longer processing times. Neither the selection nor the reordered access model is clearly and unambiguously disconfirmed by the available evidence. The selection model easily explains why context can eliminate activation of the subordinate sense but cannot prevent activation of the dominant sense. This is a nearly inevitable consequence of an autonomous lexical access system in which frequency influences lexical access. Only intralexical effects could influence activation times. Therefore only an intralexical effect could lead one sense to become available quickly enough to prevent activation of the alternative sense. In short, a subordinate sense could prevent activation of the dominant sense only if intralexical factors could speed its access sufficiently to allow it to be accepted before the dominant sense was activated. On the reordered access account, extralexical context is claimed to influence access of lexical items, i.e., its influence is not limited to selection of a single analysis or integration with preceding context. Hence on this model it is essentially an accident that a strong context favoring the subordinate sense cannot preclude activation or access of a dominant sense.

On the other side, reordered access has been claimed to provide a superior account of the Dopkins, Morris, and Rayner (1992) results, summarized in (19).

	Gaze dur on ambig.	Duration on disambig.
(19) a. context biased against dom sense:	long	med
b. context biased toward subd sense:	med	med
c. neutral context	med	long

The absence of a significant gaze duration difference on the ambiguous word in (19b) versus (19c) is taken to suggest that on some occasions only a single sense (the subordinate one) was accessed in (19b): This could explain the gaze durations on the ambiguous words as being a reflex of how many meanings are processed.

As the authors themselves noted, this argument depends crucially on interpreting a null effect. Possibly the results were also affected by the particular choice of materials, e.g., the frequent use of a pronoun in an initial subordinate clause taking the ambiguous word as antecedent. The point, however, is that at present very little, if any, evidence clearly distinguishes the interactive access account implied by the constraint-satisfaction model from (incompatible) alternative models of accessing prestored representations.

In sum, conflicting views exist concerning the correct model of accessing prestored representations. A huge literature on this issue now exists and it certainly has not been done justice here. The relevant point for present purposes is simply that questions remain about the activation and access of representations even in cases where all investigators agree that the representations are prestored.

## SUMMARY AND CONCLUSIONS

The constraint-satisfaction model claims that all analyses of a sentence are activated to some degree. Several empirical findings have been presented by proponents of the constraint-satisfaction model to support the claim that activation, or at least the comprehension difficulty of a sentence, depends on frequency and contextual constraints. In virtually every case (see above for the only exceptions), the evidence derives from comprehension times for sentences with the syntactically more complex analysis and thus may be attributed to revision times in depth-first models. The theories make diverging predictions with respect to the favored or simplest structural analysis because depth-first models predict no revision in this case and thus predict no effect of the "activation state" of the (nonlexical) syntactic or semantic analysis of more complex structures. But constraint-satisfaction predictions concerning difficulty in processing simple structures have rarely been tested; and, to date, when tested, they have not been empirically supported.

No convincing evidence has been presented suggesting the existence of a garden path due to the processor following a syntactically complex analysis in sentences with a simpler syntactic analysis. This is purely accidental on the constraint-satisfaction account since (nonstructural) frequencies and constraints favoring a syntactically complex analysis at the choice point

(e.g., the verb in a reduced relative/main clause ambiguity) should lead to choice of an incorrect analysis in sentences later disambiguated to the structurally favored analysis.

Various methodological problems limit the conclusions that can be drawn from constraint-satisfaction studies. Two problems in particular have been emphasized. First, to ask human subjects to rate entire sentence fragments from the actual experimental sentences and use this rating to predict comprehension ease, as in the typical constraint-satisfaction study, is inappropriate. Unless one is willing to claim that these particular sentence fragments are already represented in the head of the perceiver before the experiment begins, this (now popular) methodology leads to a theory of processing presupposing the very thing it should explain.

Second, no explanation is provided by the constraint-satisfaction model for what frequencies should matter. If all the action in a constraint-satisfaction model concerns contingent frequencies and probabilities, we must know how these are computed.

The constraint-satisfaction model also does not contain, utilize, or make explicit reference to global syntactic structure, making it difficult or impossible at present even to pose many central questions about syntactic processing. Because the constraint-satisfaction model stores fragments of phrase structure with particular lexical items, syntactic structure (to the extent that it is assigned at all) is lexically projected, leading to incorrect predictions in cases where lexical preferences conflict with syntactic preferences, as well as incorrect predictions about the processing of head-final phrases and about the timing and the preferences in analyzing preposed constituents. The claim that X' structure is supplied lexically, e.g., NP is stored in the lexical entry of every noun of the language, leads to difficulties in the analysis of sequences of nouns within a single NP. How are the lexically supplied NPs of all but one noun eliminated? How is the actual head noun identified? Moreover, empirical evidence supporting the minimal revisions principle suggests that the structural relation between alternative structurally complex analyses and the structurally simplest analysis influences the revision process during syntactic analysis. But in the constraint-satisfaction model the structural relation between analyses should be irrelevant at least for local ambiguities where competition between analyses is presumed to be ongoing. Hence it is difficult for the model to capture a wide array of processing preferences involving global structure, local structure, and preferences involving revisions of an already assigned analysis.

The extensive evidence suggesting the independence of grammatical structure and lexical vocabulary is left unexplained on constraint-satisfaction (and other lexical projection) views. Humans can and do structure input syntactically without any lexical heads from which to project the syntactic

structure. This should not be possible on a constraint-satisfaction view. Recent ERP evidence suggesting an independence of syntactic and nonsyntactic violations is also unexpected on the constraint-satisfaction approach.

In general the constraint-satisfaction model offers few predictions in advance and even fewer explanations. Here it has been argued that it does provide a serious if preliminary theory of accessing *prestored* representations. However, it does not offer a detailed, constrained, or satisfying account of syntactic processing. The hallmark of the human syntactic processing ability is that it is productive, able to cope with novel inputs, including those with discontinuous constituents and long-distance dependencies. Any theory of sentence processing must be able to address these issues, along with the rule-governed nature of the relation between the form of sentences and their meanings. But these interesting essential characteristics of human sentence processing are not addressed in the constraint-satisfaction model. What is offered instead is a theory of the activation of *prestored* representations without explicit detailed treatment of any other aspect of sentence processing—including those aspects which make sentence processing a distinctive and important human ability.

## REFERENCES

- Adams, B., Clifton, C., & Mitchell, D. (1995). *Lexical guidance in sentence processing: Further support for a filtering account*. Unpublished manuscript.
- Bader, M., & Lasser, I. (1994). German verb-final clauses and sentence processing: Evidence for immediate attachment. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing*. Hillsdale, NJ: Erlbaum.
- Bock, K. (1989). Closed-class immanence in sentence production. *Cognition*, 31, 163–186.
- Brown, C., Hagoort, P., & Vonk, W. (1995, March). *On-line sentence processing: Parsing preferences revealed by brain responses*. Paper presented at the Eighth Annual CUNY Sentence Processing Conference, Tucson.
- Carter, J. (1994). *Early auditory comprehension: The case for prelexical morphology and phonology*. Unpublished doctoral dissertation, University of Massachusetts.
- Clifton, C., Kennison, S., Albrecht, J. E., & Frazier, L. (1995). *Resolving the ambiguity of the word Her*. Unpublished manuscript.
- Clifton, C., Speer, S., & Abney, S. (1991). Passing arguments: Phrase structure and argument structure as determinants of initial parsing decisions. *Journal of Memory and Language*, 30, 251–271.
- Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language parsing*. Cambridge, England: Cambridge University Press.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, 19, 450–466.
- DeVincenzi, M. (1991). *Syntactic parsing strategies in Italian*. Dordrecht: Kluwer.

- DeVincenzi, M. (1995, March). *Syntactic analysis in sentence comprehension: Effects of dependency types and grammatical constraints*. Paper presented at the Eighth Annual CUNY Sentence Processing Conference, Tucson.
- Dopkins, S., Morris, R. K., & Rayner, K. (1992). Lexical ambiguity and eye fixations in reading: A test of competing models of lexical ambiguity resolution. *Journal of Memory and Language*, 31, 461-476.
- Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, 27, 429-446.
- Epstein, W. (1961). The influence of syntactical structure on learning. *American Journal of Psychology*, 74, 80-85.
- Forster, K., & Olbrei, I. (1973). Semantic heuristics and syntactic analysis. *Cognition*, 2, 319-347.
- Frazier, L. (1987a). Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory*, 5, 519-60.
- Frazier, L. (1987b). Sentence process: A tutorial review. In M. Coltheart (Ed.), *Attention and performance VII*. Hillsdale, NJ: Erlbaum.
- Frazier, L. (1990). Identifying structure under X<sup>0</sup>. In A. Jongman & A. Lahiri (Eds.), *Yearbook of Morphology*, 3, 87-109.
- Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes* 4, 93-126.
- Frazier, L., & Clifton, C. (in press). *Construal*. Cambridge, MA: MIT Press.
- Frazier, L., & Flores d'Arcais, G. B. (1989). Filler-driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language*, 28, 331-344.
- Frazier, L., & Flores d'Arcais, G. B., & Coolen, R. (1993). Processing discontinuous words: On the interface between lexical and syntactic processing. *Cognition* 47, 219-249.
- Friederici, A., Mecklinger, A., Steinhauer, K., & Hahne, A. (1995, March). *Processing violations of syntactic structure vs. violations of syntactic preferences: Evidence from ERP studies*. Paper presented at the Eighth Annual CUNY Sentence Processing Conference, Tucson.
- Garnsey, S., Pearlmutter, N., Myers, E., & MacDonald, M. C. (1995, March). *The relative contribution of verb bias and plausibility to the comprehension of temporarily ambiguous sentences*. Paper presented at the Eighth Annual CUNY Sentence Processing Conference, Tucson.
- Gibson, E., Schutze, C., & Salomon, A. (1995, March). *The relationship between the frequency and the perceived complexity of linguistic structure*. Paper presented at the Eighth Annual CUNY Sentence Processing Conference, Paper Tucson.
- Hagoort, P., & Brown, C. (1994). Brain responses to lexical ambiguity and parsing. In C. E. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing*. Hillsdale, NJ: Erlbaum.
- Hankamer, J. (1989). Morphological processing and the lexicon. In W. Marslen-Wilson (Ed.), *Lexical representation and access*. Cambridge, MA: MIT Press.
- Inoue, A., & Fodor, J. (1995). Information-paced parsing of Japanese. In R. Mazuka & N. Nagai (Eds.), *Japanese sentence processing*. Hillsdale, NJ: Erlbaum.
- Joshi, A. (1985). Processing of sentences with intransiential code-switching. In D. R. Dowty, L. Karttunen, A. M. Zwicky (Eds.), *Natural language parsing*. Cambridge, England: Cambridge University Press.
- Kennison, S. (1995) *The role of verb specific lexical information in syntactic ambiguity resolution*. Unpublished doctoral dissertation, University of Massachusetts.

- Love, T., & Swinney, D. (1995, March). *On the nature of the search in coreferential processing*. Paper presented at the Eighth Annual CUNY Sentence Processing Conference, Tucson.
- MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, 9, 157-202.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Syntactic ambiguity resolution as lexical ambiguity resolution. In C. E. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing*. Hillsdale, NJ: Erlbaum.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676-703.
- Marcus, G. F., Brinkmann, U., Clahsen, H. Wiese, R., Woest, A., & Pinker, S. (1993). German inflection: The exception that proves the rule. (Occasional Paper | 47). Cambridge, MA: MIT Center for Cognitive Science.
- Marslen-Wilson, W., & Tyler, L. (1980). On the temporal structure of spoken language comprehension. *Cognition*, 8, 1-72.
- McKinnon, R., & Osterhout, L. (1994). *Modularity in syntax reconsidered: Evidence for rapid use of constraints in sentence processing*. Manuscript submitted for publication.
- Merlo, P. (1994). A corpus-based analysis of verb continuation frequencies for syntactic processing. *Journal of Psycholinguistic Research*, 23, 435-458.
- Mitchell, D., Cuetos, F., & Corley, M. (1995, March). *The linguistic tuning hypothesis: Further corpus and experimental evidence*. Paper presented at the Eighth Annual CUNY Sentence Processing Conference, Tucson.
- Nagai, N. (1995). Constraints on topics: from a parsing perspective. In R. Mazuka & N. Nagai (Eds.), *Japanese sentence processing*. Hillsdale, NJ: Erlbaum.
- Ni, W., & Crain, S. (1990). How to resolve structural ambiguities. *Proceedings of the North Eastern Linguistic Society*, 20, 414-427.
- Obler, L., Robinson, K., Kaufman, R., & Satake, E. (1994, January). *Are suffixes read like prefixes?* Poster presented at the Annual Meeting of the Linguistic Society of America, Boston.
- Osterhout, L. (1994). Event-related brain potentials as tools for comprehending language comprehension. In C. E. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing*. Hillsdale, NJ: Erlbaum.
- Osterhout, L., & Holcomb, P. J. (1992). Event-related potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785-806.
- Pearlmutter, N. J., & MacDonald, M. C. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language*, 34, 521-542.
- Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing. *Journal of Verbal Learning and Verbal Behavior*, 22, 358-374.
- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity and lexical ambiguity. *Memory & Cognition*, 14, 191-200.
- Rayner, K., & Duffy, S. A. (1987). Eye movements and lexical ambiguity. In J. K. O'Regan & A. Levy-Schoen (Eds.), *Eye movements: From physiology to cognition*. Amsterdam: Elsevier.
- Rayner, K., & Frazier, L. (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15, 779-790.

- Schriefers, H., Friederici, A., & Kühn, K. (1995). *The processing of locally ambiguous relative clauses in German*. *Journal of Memory and Language*, 34, 499–520.
- Seidenberg, M., Tanenhaus, J., Leiman, J., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations on knowledge-based processing. *Cognitive Psychology*, 14, 489–537.
- Simpson, G. B., & Burgess, C. (1985). Activation and selection processes in the recognition of ambiguous words. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 28–39.
- Speer, S. (1995, March). *The influence of prosodic structure on the resolution of temporary syntactic clause ambiguities*. Paper presented at the Eighth Annual CUNY Sentence Processing Conference, Tucson.
- Spivey-Knowlton, M., Trueswell, J., & Tanenhaus, M. K. (1993). Context and syntactic ambiguity resolution. *Canadian Journal of Psychology*, 47, 276–309.
- Swinney, D. (1979). Lexical access during sentence comprehension: (Re-) consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18, 645–659.
- Tabossi, P., Colombo, L., & Job, R. (1987). Accessing lexical ambiguity: Effects of context and dominance. *Psychological Research*, 49, 161–167.
- Taraban, R., & McClelland, J. R. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language*, 27, 597–632.
- Traxler, M., & Pickering, M. (1995, March). *Evidence against statistical parsing*. Poster presented at the Eighth Annual CUNY Sentence Processing Conference, Tucson.
- Trueswell, J. C., & Tanenhaus, M. (1994). Toward a lexicalist framework of constraint-based syntactic ambiguity resolution. In C. E. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing*. Hillsdale, NJ: Erlbaum.
- Trueswell, J. C., Tanenhaus, M., & Garnsey, S. (1994). Semantic influences on parsing. Use of thematic role information in syntactic ambiguity resolution. *Journal of Meaning and Language*, 33, 285–318.
- Wright, B., & Garrett, M. F. (1984). Lexical decisions in sentences: Effects of syntactic structure. *Memory & Cognition*, 12, 31–45.
- Zurif, E. B., Swinney, D., Prather, P., Solomon, J., & Bushell, C. (1993). An on-line analysis of syntactic processing in Broca's and Wernicke's aphasia. *Brain and Language*, 45, 448–464.

## Exposure-Based Models of Human Parsing: Evidence for the Use of Coarse-Grained (Nonlexical) Statistical Records

Don C. Mitchell,<sup>1,4</sup> Fernando Cuetos,<sup>2</sup> Martin M. B. Corley,<sup>1</sup> and Marc Brysbaert<sup>3</sup>

Accepted August 11, 1995

*Several current models of human parsing maintain that initial structural decisions are influenced (or tuned) by the listener's or reader's prior contact with language. The precise workings of these models depend upon the "grain," or level of detail, at which previous exposures to language are analyzed and used to influence parsing decisions. Some models are premised upon the use of fine-grained records (such as lexical co-occurrence statistics). Others use coarser measures. The present paper considers the viability of models based exclusively on the use of fine-grained lexical records. The results of several studies are reviewed and the evidence suggests that, if they are to account for the data, experience-based parsers must draw upon records or representations that capture statistical regularities beyond the lexical level. This poses problems for several parsing models in the literature.*

### INTRODUCTION

There is a substantial body of evidence showing that an individual's parsing decisions are influenced in some way by his or her prior contact with com-

Aspects of this work were supported by ESRC grant No. R0023 4062 to Don Mitchell, by a Spanish Government grant DGICVT No. PB-92-0656-C04-02 to Fernando Cuetos, and by the Belgian National Fonds voor Wetenschappelijk Onderzoek, of which Marc Brysbaert is a Research Associate. We are grateful to Chuck Clifton, Barbara Hemforth, Martin Pickering, Matt Traxler, and an anonymous reviewer—all of whom made helpful comments on an earlier draft of the paper.

<sup>1</sup> University of Exeter, Devon, United Kingdom.

<sup>2</sup> University of Oviedo, Oviedo, Spain.

<sup>3</sup> University of Leuven, Leuven, Belgium.

<sup>4</sup> Address all correspondence to Don Mitchell, Department of Psychology, University of Exeter, Exeter, Devon EX4 4QG, United Kingdom.