A Computational Theory of Processing Overload and Garden-Path Effects

Edward Gibson
Department of Philosophy, Carnegie Mellon University
Pittsburgh, PA 15213-3890, USA
gibson@cs.cmu.edu

1 Introduction

The limited capacity of working memory is intrinsic to human sentence processing, and therefore must be addressed by any theory of human sentence processing. I assume that the amount of short term memory that is necessary at any stage in the parsing process is determined by the syntactic, semantic and pragmatic properties of the structure(s) that have been built up to that point in the parse. A sentence becomes unacceptable for processing reasons if the combination of these properties produces too great a load for the working memory capacity (*cf.* Frazier (1985), Gibson (1987)):

(1)

$$\sum_{i=1}^{n} A_i x_i > K$$

where:

K is the maximum allowable processing load (in processing load units or PLUs),

x_i is the number of PLUs associated with property i,

n is the number of properties,

A_i is the number of times property i appears in the structure in question.

Furthermore, I hypothesize that the human parser prefers one structure over another when the processing load (in PLUs) associated with one structure is markedly lower than the load associated with another. That is, I hypothesize there exists some arithmetic preference quantity P, corresponding to a processing load difference, such that if the processing loads associated with two representations differ by P, then only the representation associated with the smaller of the two loads survives. Given the existence of a preference quantity P, it is easy to account for garden-path effects and preferred readings of ambiguous sentences. Both effects occur because of a local ambiguity which is resolved in favor of one reading. Given two representations for the same input string that differ in processing load by at least P, only the less expensive structure will be maintained. If that structure is not compatible with the rest of the sentence and the discarded structure is part of a successful parse of the sentence, a garden-path effect results. If the parse is successful, but the discarded structure is compatible with another reading for the sentence, then only a preferred reading for the sentence has been calculated (*cf.* Gibson (1987), Gibson & Clark (1987), Clark & Gibson (1988)).[1] Thus if we know where one reading of a (temporarily) ambiguous sentence becomes the strongly preferred reading, we can write an inequality associated with this preference:

(2)

$$\sum_{i=1}^{n} A_i x_i - \sum_{i=1}^{n} B_i x_i > P$$

where:

P is the preference factor (in PLUs),

x_i is the number of PLUs associated with property i,

n is the number of properties,

A_i is the number of times property i appears in the unpreferred structure,

B_i is the number of times property i appears in the preferred structure.

In this paper I will concentrate on syntactic properties:[2] in particular, I present two properties based on the θ-Criterion and Projection Principle from Government and Binding Theory (Chomsky (1981)).[3] Once these properties are associated with processing loads, they can predict a large array of garden-path effects. Furthermore, it is demonstrated that these properties also make desirable cross-linguistic predictions with respect to unacceptability due to memory capacity overload.

The organization of this paper is given as follows. Section 2 describes the structure of the underlying parser that is assumed. Section 3 includes the proposed syntactic properties. Section 4 examines a number of locally ambiguous sentences, including some garden-paths, with respect to these properties. Section 5 explores a number of acceptable and unacceptable sentences and demonstrates that the properties proposed in Section 3 make the right predictions with respect to processing overload. Furthermore, it is demonstrated in this section that these properties seem to make the right predictions cross-linguistically. Some conclusions are given in the final section.[4]

[1] An alternative to the preference constraint presented here is the serial hypothesis, which allows at most one representation for the input string at any stage in the parse (see, for example, Frazier & Fodor (1978), Frazier (1979), Marcus (1980), Berwick & Weinberg (1984), and Pritchett (1988)). There is a longstanding debate in the psycholinguistic literature as to whether or not more than one representation for an input can be maintained in parallel. It turns out that the parallel view appears to handle some kinds of data more directly than the serial view, keeping in mind that the data are often controversial. See, for example, Kurtzman (1985) or Gorrell (1987) for a history of the debate along with evidence in support of the parallel hypothesis. Note in particular that data normally taken to be support for the serial hypothesis include garden-path effects and the existence of preferred readings of ambiguous input. However, as noted above, limiting the number of allowable representations is only one way of constraining parallelism so that these effects are also easily accounted for in a parallel framework.

[2] Note that I assume that there also exist semantic and pragmatic properties which are associated with significant processing loads.

[3] In another syntactic theory, similar properties may be obtained from the principles that correspond to the θ-Criterion and Projection Principle in that theory. For example, the completeness and coherence conditions of Lexical Functional Grammar (Bresnan (1982)) would derive properties similar to those derived from the θ-Criterion and Projection Principle. The same empirical effects should result from these two sets of properties.

[4] This paper extends work reported in Gibson (1990) by ap-

2 The Underlying Parser

The parser to which the memory limitation constraints apply must construct representations in such a way so that incomplete input will be associated with structure. Furthermore, the parsing algorithm must, in principle, allow more than one structure for an input string, so that the general constraints described in the previous section may apply to restrict the possibilities. The parsing model that I will assume is an extension of the model described in (Clark & Gibson (1988)). When a word is input to this model, representations for each of its lexical entries are built and placed in the *buffer*, a one cell data structure that holds a set of tree structures. The parsing model contains a second data structure, the *stack-set*, which contains a set of stacks of buffer cells. The parser builds trees in parallel based on possible attachments made between the buffer and the top of each stack in the stack-set. The buffer and stack-set are formally defined in (3) and (4).

(3) A *buffer cell* is a set of structures $\{ S_1, S_2, ..., S_n \}$, where each S_i represents the same segment of the input string. The *buffer* contains one buffer cell.

(4) The *stack-set* is a set of stacks of buffer cells, where each stack represents the same segment of the input string:

$$\{ (\{ S_{1,1,1}, S_{1,1,2}, ..., S_{1,1,n_{1,1}} \}, \{ S_{1,m_1,1}, S_{1,m_1,2}, ..., S_{1,m_1,n_{1,m_1}} \}) \; ... \; (\{ S_{p,1,1}, S_{p,1,2}, ..., S_{p,1,n_{p,1}} \}, \{ S_{p,m_p,1}, S_{p,m_p,2}, ..., S_{p,m_p,n_{p,m_p}} \}) \}$$

where:

p is the number of stacks;

m_i is the number of buffer cells in stack i;

and $n_{i,j}$ is the number of tree structures in the jth buffer cell of stack i.

The motivation for these data structures is given by the desire for a completely unconstrained parsing algorithm upon which constraints may be placed. This algorithm should allow all possible parser operations to occur at each parse state. There are only two parser operations: attaching a node to another node and pushing a buffer cell onto a stack. In order to allow both of these operations to be performed in parallel, it is necessary to have the given data structures: the buffer and the stack-set.

2.1 Node Attachment

I assume a hypothesis-driven node projection algorithm of the following form. In order to satisfy $\overline{X}$ Theory (Jackendoff (1977), Chomsky (1986b)), a maximal projection is constructed for each of a word's lexical entries when that word is input. For each of these structures, the lexical requirements and category of the structure causes the local prediction to the right of further categories. These predicted structures are called *hypothesized* nodes or H-nodes. All other structures are called *confirmed* nodes or C-nodes. For example, when the noun *dog* is input, a confirmed noun phrase and a hypothesized clausal phrase are among the structures built:[5]

(5) a. $[_{NP} [_{N'} [_N \text{dog}]]]$
b. $[_{XP_{clause}} [_{NP} [_{N'} [_N \text{dog}]]] [_{X'_{clause}} [_{X_{clause}} e]]]$

Node attachment in this framework consists of matching hypothesized nodes on top of a stack in the stack-set against nodes in the buffer. If the features of two such nodes are compatible, then an attachment takes place, the result being the unification of the stack and buffer nodes. Since attachment always involves matching an H-node, all obligatory arguments are projected as H-nodes.

3 Dynamic Application of the θ-Criterion and Projection Principle

Following much current work in syntactic theory, I assume that the grammar consists of a set of environments where properties such as thematic role and Case may be assigned, along with a set of filters, each of which rules out representations lacking a necessary property. This approach to syntactic theory has been labeled the Principles and Parameters Approach (PPA) (Chomsky (1986a)). The particular syntactic theory that I will assume is known as Government and Binding theory (see Chomsky, (1981, 1986a, 1986b) along with the references cited in each), although the methodology of this work will apply to any syntactic theory of this form.

A filter-based syntactic theory presents an obvious set of candidates for load-bearing structural properties: the set of local violations of all syntactic filters. That is, given a constraint-based syntactic theory, it is reasonable to assume that there is a processing load associated with the local violation of each syntactic filter. In particular, I will consider the θ-Criterion and Projection Principle from Government and Binding Theory with respect to a theory of processing. These principles are given in (6) and (7):

(6) The Projection Principle:
Lexical requirements must be satisfied at all levels of representation. (paraphrased from Chomsky (1981) p. 29).[6]

(7) The θ-Criterion:
Each argument bears one and only one θ-role (thematic role) and each θ-role is assigned to one and only one argument (Chomsky (1981) p. 36).

Note that the second part of the θ-Criterion– that each θ-role be assigned– follows from the Projection Principle. Thus the θ-Criterion that I will assume consists only of the first clause of (7):

(8) The θ-Criterion (simplified):
Each argument bears one and only one θ-role.

The dynamically applied θ-Criterion can be stated as the following processing property:

(9) The Property of Thematic Reception (PTR):
Associate a load of x_{TR} PLUs of short term memory to each constituent that is in a position that can receive a thematic role in some co-existing structure, but whose θ-assigner is not unambiguously identifiable in the structure in question.

plying the methodology described here to the Projection Principle as well as the θ-Criterion. While the effects reported in the earlier paper still hold, many additional results are obtained and reported here.

[5]A noun phrase is projected to an H-node clausal (or predicate) phrase since nouns may be the subjects of predicates.

[6]Government and Binding Theory assumes the existence of a number of levels of representation. I assume that the level most relevant to parsing is surface structure or S-structure. Thus the Projection Principle applied to S-structure dictates that lexical requirements be satisfied at that level.

The dynamically applied Projection Principle gives a similar property. This property is stated in terms of *thematic* elements. Following early work in linguistic theory, I distinguish two kinds of categories: *functional* categories and *thematic* or *content* categories (see, for example, Fukui and Speas (1986) and Abney (1987) and the references cited in each). Thematic categories include nouns, verbs, adjectives and prepositions; functional categories include determiners, complementizers, and inflection markers. I hypothesize that thematic elements are more visible to the parser than their functional counterparts. This assumption is made explicit in the Property of Lexical Requirement, the dynamic version of the Projection Principle:

(10) The Property of Lexical Requirement (PLR):
Associate a load of x_{LR} PLUs of short term memory to each lexical requirement that is satisfied by a hypothesized constituent containing no thematic elements.

Note that all lexical requirements minimally involve the existence of a hypothesized structure. Thus the Property of Lexical Requirement ignores those structures whose lexical requirements are satisfied by either confirmed nodes or hypothesized nodes that contain thematic elements. The PLR will penalize only those structures with unsatisfied lexical requirements, where *unsatisfied requirements* consist of thematic element-less hypothesized structures.

When particular processing loads are associated with each of these properties, they will make a large number of empirical predictions in the theory of overload and preference. Since both the PTR and the PLR are properties dealing with the interpretation of an utterance, it is reasonable to assume as a default that the loads associated with these two properties, x_{TR} PLUs and x_{LR} PLUs respectively, are the same.[7]

(11) $x_{TR} = x_{LR} = x_{Int}$

It turns out that this assumption is consistent with all inequalities that are obtained in this paper.

4 Ambiguity and the Properties of Thematic Reception and Lexical Requirement

In order to determine what load is associated with each of the Properties of Thematic Reception and Lexical Requirement, I will first examine locally ambiguous sentences that either cause or do not cause garden-path effects. Consider sentence (12) with respect to the PTR and PLR:

(12) John expected Mary to like Fred.

The verb *expect* is ambiguous: either taking an NP complement as in the sentence *John expected Mary* or taking an IP complement as in (12).[8] Thus there is a local ambiguity in (12) at the point of parsing the NP *Mary*. Despite this local ambiguity, there is no difficulty parsing (12). Consider the state of the parse of (12) after the word *Mary* has been processed:

(13) a. $[_{IP}$ $[_{NP}$ John $]$ $[_{VP}$ expected $[_{NP}$ Mary $]]]$
b. $[_{IP}$ $[_{NP}$ John $]$ $[_{VP}$ expected $[_{IP}$ $[_{NP}$ Mary $]$ $]]]$

In (13a) the NP *Mary* is attached as the NP complement of *expected*. In this representation there is no load associated with either of the Properties of Thematic Reception or Lexical Requirement since all constituents that are positions to receive thematic roles, do so, and all lexical requirements are satisfied. In (13b) the NP *Mary* is the specifier of a hypothesized IP node which is attached as the complement of the other reading of *expected*. This representation is associated with at least x_{TR} PLUs (= x_{Int} PLUs) since the NP *Mary* is in a position that can be associated with a thematic role (the subject position), but does not yet receive one in this structure. No load is associated with the Property of Lexical Requirement, however, since the lexical requirements of the verb *expected* are satisfied by nodes that contain thematic elements. Since there is no difficulty in processing sentence (12), the load difference between these two structures cannot be greater than P PLUs, the preference factor assumed in inequality (2). Thus the inequality in (14) is obtained:

(14) $x_{Int} \leq P$

Since the load difference between the two structures is not sufficient to cause a strong preference, both structures are maintained. Note that this is a crucial difference between the theory presented here and the theory presented in Frazier & Fodor (1978), Frazier (1979) and Pritchett (1988). In each of these theories, only one representation can be maintained, so that either (13a) or (13b) would be preferred at this point. In order to account for the lack of difficulty in parsing (12), Pritchett assumes that backtracking in certain situations is not expensive. No such stipulation is necessary in the framework given here.

Now consider a second locally ambiguous sentence, one that results in a garden-path effect:[9]

(15) # I put the candy on the table in my mouth.

Sentence (15) is locally ambiguous at the point of parsing the word *on*. This preposition may attach as either an argument of the verb *put*, or as a modifier of the noun *table*. The argument attachment is locally preferred, although it turns out that this attachment is not compatible with the rest of the sentence. Thus a garden-path effect results. In order to see how the Properties of Thematic Reception and Lexical Requirement can account for this garden-path effect, consider the state of the parse after the word *on* has been input:

(16) a. $[_{IP}$ $[_{NP}$ I $]$ $[_{VP}$ $[_{V'}$ $[_{V}$ put $]$ $[_{NP}$ the candy $]$ $[_{PP}$ on $[_{NP}$ $]]$ $]]]$
b. $[_{IP}$ $[_{NP}$ I $]$ $[_{VP}$ $[_{V'}$ $[_{V}$ put $]$ $[_{NP}$ the candy $[_{PP}$ on $[_{NP}$ $]]]$ $]]]$

The load associated with structure (16a) is x_{LR} PLUs since, although the lexical requirements of the verb *put* are satisfied, the lexical requirements of the preposition *on* remain unsatisfied. On the other hand, the load associated with the modifier attachment is $2x_{LR} + x_{TR}$ PLUs since 1) both the verb *put* and the preposition *on* have unsatisfied

[7] In fact, both the θ-Criterion and the Projection Principle are generally believed to follow from a more general principle: that of Full Interpretation (Chomsky (1986a)). If this is so, then the PLR and the PTR reduce to a single property: that of local uninterpretability. However, the principle of Full Interpretation has not yet been adequately formalized. Thus I will continue to appeal only to its components.

[8] Following current notation in GB Theory, IP=S and CP=S′ (Chomsky (1986b)).

[9] I will prefix sentences that are difficult to parse because of memory limitations with the symbol "#". Hence sentences that are unacceptable due to processing overload will be prefixed with "#", as will be garden-path sentences.

lexical requirements and 2) the PP headed by *on* receives a thematic role in the argument attachment structure, while it receives no such role in the current structure. Thus the difference between the loads associated with the two structures is $x_{LR} + x_{TR}$ PLUs = $2x_{Int}$ PLUs. Since the argument attachment structure is strongly preferred over the other structure, I hypothesize that this load is greater than P PLUs:

(17) $2x_{Int} > P$

Now consider the well-known garden-path sentence in (18) (Bever (1970)):

(18) # The horse raced past the barn fell.

The structure for the input *the horse raced* is ambiguous between at least the two structures in (19):

(19) a. $[_{IP}$ $[_{NP}$ the horse $]$ $[_{VP}$ raced $]]$
b. $[_{IP}$ $[_{NP}$ the $[_{N'}$ $[_{N'}$ horse$_i$ $]$ $[_{CP}$ O_i raced $]$ $]]$ $]$

Structure (19a) has no load associated with it due to either the PLR or the PTR. Crucially note that the verb *raced* has an intransitive reading so that no load is required via the Property of Lexical Requirement. On the other hand, structure (19b) requires a load of $2x_{TR}$ PLUs since 1) the noun phrase *the horse* is in a position that can receive a thematic role, but currently does not and 2) the operator O_i is in a position that may be associated with a thematic role, but is not yet associated with one.[10] Thus the difference between the processing loads associated with structures (19a) and (19b) is $2x_{TR}$ PLUs = $2x_{Int}$ PLUs. By the inequality in (17), this difference is sufficient to cause the preference of the less expensive structure. Hence the garden-path effect in (18) is predicted.

Consider (20), a sentence whose structure and local ambiguities are very similar to those in (18):

(20) The bird found in the room was dead.

Although the structures and local ambiguities in (20) and (18) are similar, (18) causes a garden-path effect while, surprisingly, (20) does not. To determine why (20) is not a garden-path sentence we need to examine the local ambiguity when the word *found* is read:

(21) a. $[_{IP}$ $[_{NP}$ the bird $]$ $[_{VP}$ $[_{V'}$ $[_{V}$ found $[_{NP}$ $]$ $]]]]$
b. $[_{IP}$ $[_{NP}$ the $[_{N'}$ $[_{N'}$ bird$_i$ $]$ $[_{CP}$ O_i found $]$ $]]$ $]$

The crucial difference between the verb *found* and the verb *raced* is that *found* obligatorily requires a noun phrase object, while *raced* does not. Since the lexical requirements of the verb *found* are not yet satisfied in structure (21a), this representation is associated with x_{LR} PLUs of memory load. Like structure (19b), structure (21b) requires $2x_{TR}$ PLUs. Thus the difference between the processing loads of structures (21a) and (21b) is $2x_{TR} - x_{LR}$ PLUs = x_{Int} PLUs. By the result obtained from sentence (12), this load difference is not sufficient to force a strong preference. Thus the lack of garden-path effect in (20) is explained. Furthermore, these results correctly predict that sentence (22) is not a garden-path sentence either:

(22) The bird found in the room enough debris to build a nest.

Consider now (23):

(23) # I believe that John smokes annoys Mary.

[10] In fact, this operator will be associated with a thematic role as soon as a gap-positing algorithm links it with the object of the passive participle *raced*. However, when the attachment is initially made, no such link yet exists: the operator will initially be unassociated with a thematic role.

When the complementizer *that* is input, it can attach as either the argument of the verb *believe* or as subject of *believe*'s complement clause. The argument attachment is strongly preferred and a garden-path effect results in (23), since it is the other reading that is necessary for a successful parse. Consider the state of the parse after the word *that* has been input:

(24) a. $[_{IP}$ $[_{NP}$ I $]$ $[_{VP}$ believe $[_{CP}$ $[_{C'}$ $[_{C}$ that $[_{IP}$ $]$ $]]$ $]]]$
b. $[_{IP}$ $[_{NP}$ I $]$ $[_{VP}$ believe $[_{CP}$ $[_{C'}$ $[_{C}$ e $]$ $[_{IP}$ $[_{CP}$ that $[_{IP}$ $]]]$ $]]$ $]]$

Consider first structure (24a). All positions that can receive thematic roles, do so. Thus there is no load associated with (24a) with respect to the Property of Thematic Reception. However the complementizer *that* requires an IP complement, so that x_{LR} PLUs are associated with (24a). Consider now (24b). First of all, the CP headed by *that* is in thematic role receiving position, but it does not yet receive a thematic role. Thus (24b) is associated with at least x_{TR} PLUs. Furthermore, both the lexical complementizer *that* and the argument non-lexical complementizer of *believe* have lexical requirements that must be satisfied, but as yet are unsatisfied. Thus structure (24b) is associated with an additional $2x_{LR}$ PLUs via the Property of Lexical Requirement. Thus the total load associated with structure (24b) is $x_{TR} + 2x_{LR}$ PLUs. Hence the difference between the loads of the structures in (24) is $x_{TR} + x_{LR}$ PLUs = $2x_{Int}$ PLUs. As we have seen earlier, this load difference is sufficient for a preference to occur. Thus the garden-path effect in (23) is predicted, as desired.

See Gibson (1990) for explanations of further garden-path effects inside a similar framework.

5 Processing Overload and the Properties of Thematic Reception and Lexical Requirement

The Properties of Thematic Reception and Lexical Requirement also give a plausible account of unacceptability due to processing overload. Recall that I assume that a sentence is unacceptable because of short term memory overload if the combination of memory associated with properties of the structures built at some stage of the parse of the sentence is greater than the allowable processing load K. Consider the unacceptable center-embedding sentences in (25):

(25) a. # The man that the woman that won the race likes eats fish.
b. # The man that the woman that the dog bit likes eats fish.

Consider one structure that results from parsing either of the sentences in (25) after the second complementizer *that* has been input:

(26) $[_{IP}$ $[_{NP}$ the $[_{N'}$ $[_{N'}$ man$_i$ $]$ $[_{CP}$ $[_{NP}$ O_i $]$ that $[_{IP}$ $[_{NP}$ the $[_{N'}$ $[_{N'}$ woman$_j$ $]$ $[_{CP}$ $[_{NP}$ O_j $]$ that $[_{IP}$ $]]$ $]]]]$ $]]]$

First consider (26) with respect to the Property of Thematic Reception. There are two lexical noun phrases in (26) that need thematic roles but lack them. Furthermore, there are two non-lexical NPs, operators, that are in positions that may prospectively be linked to thematic roles. Thus the load associated with (26) is at least $4x_{TR}$ PLUS. Now consider (26) with respect to the Property of Lexical Requirement. Only the two complementizers have lexical requirements in (26), and only the most recent of these is unsatisfied, since the lexical requirements of the first are satisfied by a hypothesized node with thematic content. Thus the total load associated with (26) is $4x_{TR}+x_{LR}$ PLUS $= 5x_{Int}$ PLUs. I hypothesize that this load is too much for the limited capacity of working memory:

Indeed, when noun phrases with two levels of center-embedded relative clauses appear post-verbally, the results are still unacceptable, although perhaps better:

(27) a. ?# I saw the man that the woman that won the race likes.
  b. # I saw the man that the woman that the dog bit likes.

Since the NP *the man* receives a thematic role as soon as it is parsed, it does not contribute to the processing load in either of the sentences in (27). However, other factors in determining the maximal processing load of the sentences in (27) remain the same. Thus the maximal load in each of (27a) and (27b) is $4x_{Int}$ PLUs. Since each of these sentences is unacceptable, I hypothesize that this load is more than can be handled by the short term memory capacity:

(28) $4x_{Int} > K$

Note that sentences with only one relative clause modifying a subject NP are acceptable, as is exemplified in (29)

(29) The man that Mary likes eats fish.

Since (29) is acceptable, its load is below the maximum at all stages of its processing. Under the assumptions presented in this paper, the processing load associated with (29) will be greatest when the complementizer *that* is input. At this point in the parse, there will be one lexical NP, *the man*, and one non-lexical NP, an operator, that require thematic roles but currently lack them. Furthermore, the complementizer *that* requires a complement IP which is not yet present. Thus the total load associated with (29) at the point of parsing *that* is $3x_{Int}$ PLUs.[11] Since there is no difficulty in parsing (29), we arrive at the inequality in (30):

(30) $3x_{Int} \leq K$

Thus I assume that the maximum processing load that people can handle lies above $3x_{Int}$ PLUs but below $4x_{Int}$ PLUs. Further data support this conclusion. For example, consider the contrast between the sentences in (31b):

(31) a. That John smokes annoys me.
  b. # That for John to smoke would annoy me is obvious.

Although it is possible for a clause to be subject of a matrix clause, as in (31a), an unacceptable sentence results when the subject clause contains a further clause as its subject, as in (31b). The acceptability of (31a) together with the unacceptability of (31b) can be easily explained in the framework offered here. Consider first sentence (31a). The maximal processing load associated with the parse of (31a) occurs as the words *that* and *John* are processed. In both of these states the load is $2x_{Int}$ PLUs, less than the available memory capacity. Thus there is no difficulty in the processing of (31a). Consider, however, the state of the parse of (31b) after the complementizer *for* has been input:

(32) $[_{IP}$ $[_{CP}$ that $[_{IP}$ $[_{CP}$ for $[_{IP}$]]]]]

There are two complementizer phrases, both in thematic positions, which currently lack thematic roles. Furthermore, both complementizers have lexical requirements that are currently unsatisfied: that is, the complement of each complementizer neither contains a thematic element nor is a confirmed node. Thus the total load associated with (32) is $4x_{Int}$ PLUs, which is enough to force processing overload. Thus the unacceptability of (32) is explained.

Furthermore, the acceptability of (33) comes as no surprise to the account presented here:

(33) I believe that for John to smoke would annoy me.

In contrast to (31b), the first complementizer in (33) receives a thematic role as soon as it is processed. Thus the maximal processing load associated with the parse of (33) is only $3x_{Int}$ PLUs, not enough to overload short term memory.

5.1 Processing Overload: Cross-Linguistic Predictions

The examples discussed thus far are all English ones. A strong test of the theory presented here is presented by data from other languages. First let us consider center-embedded relative clauses in languages closely related to English. In particular, consider Dutch and German. It turns out that multiply center-embedded relative clauses become difficult in these languages at the same point as they do in English: on the second embedding. For example, the German sentence (34) is unacceptable, as expected:

(34) # Der Mann den die Frau die der Hund biß sah schwam.
"The man that the woman that the dog bit saw swam."

Unlike English, however, German and Dutch are verb final in subordinate clauses, so that verbs with lexical requirements for three thematic elements pose an interesting test to the theory. If the theory presented here is correct and it generalizes cross-linguistically, then constructions with three initial θ-role-requiring constituents should be acceptable. It turns out that this is in fact the case, as the German example (35) illustrates:

(35) Ich glaube, daß John Mary das Geschenk gegeben hat.
I believe that John Mary the present given has
"I believe that John has given Mary the present."

After the word *Geschenk*, there are three noun phrases that require thematic roles, but currently lack them. All lexical requirements are satisfied at this point in the parse, so the total load associated with this parse state is $3x_{Int}$ PLUs. Thus (35) is predicted to acceptable, as desired.

Another good test for the theory presented here comes from cross-serial dependencies in Dutch.[12] In examples of cross-serial dependency, noun phrase arguments appear at the beginning of a subordinate clause, followed by their thematic role assigning verbs. It turns out that

[11] In fact, the load remains at $3x_{Int}$ PLUs when the NP *Mary* is input: the NP *Mary* requires a thematic role, thus adding to the processing load, but the lexical requirements of the complementizer *that* also become satisfied at this point, since a thematic element, *Mary*, is now present in the hypothesized IP complement.

[12] See Bresnan *et al* (1982) for a discussion of the syntax of such constructions.

constructions with three initial noun phrases are perfectly acceptable, as is exemplified in (36):

(36) ... dat Jan Piet de kinderen zag helpen zwemmen.
... that Jan Piet the children saw help swim
"... that Jan saw Piet help the children swim."

However, these constructions lose their acceptability with the addition of further NP arguments:

(37) a. ?# ... dat Jan Piet Marie de kinderen zag helpen laten zwemmen.
... that Jan Piet Marie the children saw help make swim
"... that Jan saw Piet help Marie make the children swim."

b. # ... dat Jan Piet Marie Karel de kinderen zag helpen laten leren zwemmen.
... that Jan Piet Marie the children saw help make teach swim
"... that Jan saw Piet help Marie make Karel teach the children to swim."

This result is predicted in the framework presented here. Four NP arguments locally lacking thematic roles force a load of $4x_{Int}$, too much for human short term memory capacity.

Evidence from the processing of Japanese also supports the memory capacity results obtained here. Japanese is a verb final language, so that subjects and objects appear before the verb. Verbs that take clausal complements provide an interesting test case for the theory presented here, since it is grammatical to place all NP arguments before the thematic role assigning verbs. For example, consider (38):

(38) Jon wa Fred ga Biru wo sukida to omotteiru.
John TOPIC Fred NOM Bill ACC likes COMP thinks
"John thinks that Fred likes Bill"

Sentence (38) is perfectly acceptable, as is predicted by the theory presented here. The processing load associated with (38) peaks at the point that the NP *Biru* is input: at this point there are three NP arguments which require thematic roles but currently lack them. As a result, the processing load associated with this processing state is $3x_{Int}$ PLUs, not enough to cause overload. However, when more than three NP arguments appear sentence initially, acceptability is lost, as predicted by the processing overload results obtained here:

(39) a. ?# Jon wa Mary ga Fred ga Biru wo sukida to sinjiteiru to omotteiru.
John TOPIC Mary NOM Fred NOM Bill ACC likes COMP believes COMP thinks
"John thinks that Mary believes that Fred likes Bill."

b. # Jon wa Mary ga Fred ga Sam ga Biru wo sukida to omotteiru to sinjiteiru to omotteiru.
John TOPIC Mary NOM Fred NOM Sam NOM Bill ACC likes COMP thinks COMP believes COMP thinks
"John thinks that Mary believes that Fred thinks that Sam likes Bill."

6 Conclusions

Since the structural properties that are used in the formation of the inequalities are independently motivated, and the system of inequalities is solvable, the theory of human sentence processing presented here makes strong, testable predictions with respect to the processability of a given sentence. Furthermore, the success of the method provides empirical support for the particular properties used in the formation of the inequalities. Thus a theory of PLUs, the preference factor P and the overload factor K provides a unified account of 1) acceptability and relative acceptability; 2) garden-path effects; and 3) preferred readings for ambiguous input and perception of ambiguity.

7 Acknowledgements

I would like to thank my informants: Alex Franz, Hiroaki Kitano, Ingrid Meyer, Teruko Mitamura and Michel Nederlof. I would also like to thank Robin Clark, Dan Everett, Rick Kazman, and Eric Nyberg for comments on earlier drafts of this work. All remaining errors are my own.

8 References

Abney (1987) *The English Noun Phrase in its Sentential Aspect*, MIT Ph.D. dissertation, Cambridge, MA.

Berwick, R. & Weinberg, A. (1984) *The Grammatical Basis for Linguistic Performance*, MIT Press, Cambridge, MA.

Bever (1970) "The Cognitive Basis for Linguistic Structures," in J.R. Hayes (ed.), *Cognition and the Development of Language*, Wiley, New York, NY.

Bresnan, J. (1982) *The Mental Representation of Grammatical Relations*, MIT Press, Cambridge, MA.

Bresnan, J., Kaplan, R.M., Peters, S., & Zaenen, A. (1982) "Cross-serial Dependencies in Dutch," *Linguistic Inquiry* 13:613-636.

Chomsky, N. (1981) *Lectures on Government and Binding*, Foris, Dordrecht, The Netherlands.

Chomsky, N. (1986a) *Knowledge of Language: Its Nature, Origin and Use*, Praeger Publishers, New York, NY.

Chomsky, N. (1986b) *Barriers*, Linguistic Inquiry Monograph 13, MIT Press, Cambridge, MA.

Clark, R. & Gibson, E. (1988) "A Parallel Model for Adult Sentence Processing," *Proceedings of the Tenth Cognitive Science Conference*, McGill University, Montreal, Quebec.

Frazier, L., & Fodor, J.D. (1978) "The Sausage Machine: A New Two-stage Parsing Model," *Cognition* 6:291-325.

Frazier, L. (1979) *On Comprehending Sentences: Syntactic Parsing Strategies*, University of Massachusetts Ph.D. dissertation.

Frazier, L. (1985) "Syntactic Complexity," in D. Dowty, L. Karttunen, & A. Zwicky (eds.), *Natural Language Processing: Psychological, Computational and Theoretical Perspectives*, Cambridge University Press, Cambridge, U.K.

Frazier, L. (1987) "Syntactic Processing Evidence from Dutch," *Natural Language and Linguistic Theory* 5:519-559.

Fukui, N., & Speas, M. (1986) "Specifiers and Projections," *MIT Working Papers in Linguistics* 8, Cambridge, MA.

Gibson, E. (1987) "Garden-Path Effects in a Parser with Parallel Architecture," *Proceedings of the Fourth Eastern States Conference on Linguistics*, The Ohio State University, Columbus, OH.

Gibson, E. (1989) "Parsing with Principles: Predicting a Phrasal Node Before Its Head Appears," *Proceedings of the First International Workshop on Parsing Technologies*, Carnegie Mellon University, 1989.

Gibson, E. (1990) "Memory Capacity and Sentence Processing," *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, Pittsburgh, PA.

Gibson, E., & Clark, R. (1987) "Positing Gaps in a Parallel Parser," *Proceedings of the Eighteenth North East Linguistic Society Conference*, University of Toronto, Toronto, Ontario.

Gorrell, P.G. (1987) *Studies of Human Syntactic Processing: Ranked-Parallel versus Serial Models*, University of Connecticut Ph.D. dissertation.

Jackendoff, R. (1977) *X-bar Syntax: A Study of Phrase Structure*, Linguistic Inquiry Monograph 2, MIT Press, Cambridge, MA.

Kurtzman, H. (1985) *Studies in Syntactic Ambiguity Resolution*, MIT Ph.D. dissertation.

Marcus, M. (1980) *A Theory of Syntactic Recognition for Natural Language*, MIT Press, Cambridge, MA.

Pritchett, B. (1988) "Garden Path Phenomena and the Grammatical Basis of Language Processing," *Language* 64:539-576.