**Carnegie Mellon University**

# A Computational Theory of Human Linguistic Processing: Memory Limitations and Processing Breakdown

**Edward Albert Fletcher Gibson**

A Dissertation
Submitted to the Graduate School
in Partial Fulfillment of the Requirements
for the Degree
Doctor of Philosophy
in
Computational Linguistics

Department of Philosophy
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213
May 1991

Thesis Committee:

Robin Clark (co-chairman), Department of Linguistics, University of Geneva
Dan Everett (co-chairman), Department of Linguistics, University of Pittsburgh
Jaime Carbonell, School of Computer Science, Carnegie Mellon University
Howard Kurtzman, Department of Psychology, Cornell University
Carl Pollard, Department of Linguistics, Ohio State University
Masaru Tomita, School of Computer Science, Carnegie Mellon University

## Abstract

This thesis gives a theory of sentence comprehension that attempts to explain a number of linguistic performance effects, including garden-path effects, preferred readings for ambiguous input and processing overload effects. It is hypothesized that the human parser heuristically determines its options based upon evaluation of possible representations with respect to lexical, syntactic, semantic and pragmatic properties, each of which is associated with a weight. Processing overload effects are explained by the assumption of the existence of a maximum load corresponding to the limited capacity of short term memory: a structure becomes unacceptable at a particular parse state if the combination of the processing weights associated with its properties at that state is greater than the available capacity.

Furthermore, it is assumed that the language processor is an automatic device that maintains only the best of the set of all compatible representations for an input string. This thesis assumes a formulation of representational evaluation within a parallel framework: one structure is preferred over another if the processing load associated with the first structure is markedly lower than the processing load associated with the second. Thus a garden-path effect results if the unpreferred structure is necessary for a successful parse of the input.

Four properties of linguistic representations are presented within this framework. The first two – the Properties of Thematic Reception and Transmission – derivable from the $\theta$-Criterion from Government-Binding (GB) Theory (Chomsky (1981)); the third – the Property of Lexical Requirement – derivable from the Projection Principle of GB Theory; and the fourth – the Property of Recency Preference – prefers local attachments over more distant attachments (*cf.* Kimball (1973), Frazier (1979)). This thesis shows how these properties interact to give a partially unified theory of many performance effects.

**Acknowledgements**

I would like to thank the members of my committee – Robin Clark, Dan Everett, Carl Pollard, Jaime Carbonell, Howard Kurtzman and Masaru Tomita – for their assistance and advice during this research. I would especially like to thank my co-chairmen Dan Everett and Robin Clark. Both Dan and Robin have given exceptional guidance and encouragement during the last five years. I feel lucky to have been able to work with each of them. I would also like to thank the other members of my thesis committee for being so careful and thorough. I hope that I can do as good a job for others sometime in the future.

The graduate students in the Computational Linguistics program at CMU and the University of Pittsburgh have also contributed a great deal to my life as a graduate student. In particular, I owe most to my 1986 CL program classmates Rick Kazman and Eric Nyberg. The three of us worked well together while we were here. I hope that we can continue to work together when we move on to different locations. I would also like to thank other students in the CL program that contributed significantly to my experience in Pittsburgh: Kathy Baker, Alex Franz, Prahlad Gupta, Boyan Onyshkevych and Margalit Zabludowski.

Additional thanks go to David Evans, Rich and Sally Thomason, Brad Pritchett, Teddy Seidenfeld, Bob Carpenter and the other faculty members in the CL program. Furthermore, I would like to thank Renee Schafer and Jackie DeFazio of the Department of Philosophy, as well as Radha Rao and Barbara Moore of the Center for Machine Translation for giving me excellent administrative support. I must also acknowledge the people at the Laboratory for Computational Linguistics, where I worked from 1986-1988, and the Center for Machine Translation, where I worked from 1989-1991. I have benefited greatly from interactions at the CMT with Teruko Mitamura, Ralf Brown and Ingrid Meyer. I would especially like to thank Sergei Nirenburg and Lori Levin for their excellent advice on many issues over the past two years.

I also want to express my gratitude to my friends David, Andrea, Rick and Nancy for invaluable discussions on sometimes less scientific issues away from the terminal.

Finally and most importantly I want to thank my wife Sarah for all her love, support, and for doing all the figures in this thesis.

This thesis is dedicated to my mother Elizabeth and to the memory of my father, James Douglas Gibson, 1909-1988.

**Contents**

## List of Figures

## 1. Introduction

A longstanding topic of research in linguistics and psychology is determining how people process language. That is, given 1) linguistic knowledge made up of lexical, morphological, phonological, syntactic, and semantic information and 2) non-linguistic knowledge consisting of pragmatic and contextual information, how does a human being understand linguistic input? What algorithms are used and what constraints, if any, apply to these algorithms?

Computer scientists are also interested in the answers to these questions because of the possible applications such an algorithm might have for automatic natural language processing. For example, computer scientists who are interested in parsing want to obtain representations for input natural language as fast as possible. One way to attack such a problem is to devise efficient algorithms for obtaining all possible parses for an input string (see *e.g.*, Aho & Ullman (1972) and Tomita (1987) and the references cited in each). Another attack on the problem relies on the fact that natural languages, unlike programming languages, are extremely ambiguous. For example, a prepositional phrase that occurs late in a sentence can attach to many of the verb phrases and noun phrases that precede it. Consider the following English example:

(1) The man saw the boy on the hill with the telescope.

The prepositional phrase (PP) *on the hill* may attach to either the noun phrase (NP) *the hill* or to the verb phrase (VP) headed by *saw*. Furthermore, the PP *with the telescope* can attach to the NP *the hill*, the NP *the boy* or the VP headed by *saw*. The combination of all of these possibilities gives five possible structures for (1). In general, the number of parses with $n$ prepositional phrases following a NP V NP sequence is the Catalan number of $n+1$. Thus long sentences may have hundreds or even thousands of parses.

Hence obtaining all possible parses of a natural language sentence may be computationally expensive. Furthermore, calculating all of these structures is wasteful, since most of the logically possible structures for an input sentence go unnoticed by people. For example, although (1) is ambiguous, there are readings which people strongly prefer over others. In particular, the preferred interpretation of (1) in the null context links the PP *with the telescope* to the verb *saw* rather than to either of its preceding NPs. As a result of this observation, it is reasonable to investigate how people parse sentences in order to see how a practical NLP system might selectively ignore many ambiguities and therefore be more efficient.

Consider the following simple sentence along with a plausible candidate for its syntactic structure with respect to the question of how people compute linguistic structure:

(2) a. John saw Mary.
b. $[_S\ [_{NP}$ John $]\ [_{VP}\ [_V$ saw $]\ [_{NP}$ Mary $]]]]$

Assuming that (2b) is the syntactic structure that English speakers build for (2a), how is it that this structure is obtained from (2a)? Many different parsing algorithms have been proposed in the psychological and computer science literature. The psychological plausibility of an algorithm depends on two factors, the first being its ability to account for psychological effects. For example, an algorithm that parses (2a) from right to left would not satisfy the first criterion, since English speakers read text from left to right. Similarly, a parsing algorithm that requires that a complete input sentence be present before parsing can take place is not psychologically plausible, since it is well established that people process linguistic input incrementally (Marslen-Wilson (1975, 1987); Tyler & Marslen-Wilson (1977); Swinney (1979); Marslen-Wilson & Tyler (1980); Shillcock (1982); Garrod & Sanford (1985); Tanenhaus, Carlson & Seidenberg (1985) among others).

The second factor contributing to a parsing algorithm's psychological plausibility is that of simplicity as determined by Occam's razor. That is, one algorithm is more psychologically plausible than another if it includes fewer stipulations, other factors being equal. For example, suppose that one parsing algorithm performs all the steps in the same order as another algorithm. Furthermore, suppose that the first of these algorithms includes a step that the second does not, such that this step takes no time and has no empirical consequences. The two parsing algorithms make exactly the same empirical predictions, but the first algorithm is preferred on conceptual grounds: it is more parsimonious.

This thesis seeks to provide a computational model of human sentence processing, paying particular attention to evidence from linguistic input that causes processing breakdown. Since a crucial part of this research involves investigation of processing difficulty, it is important to distinguish knowledge of language from the use of that knowledge so that processing difficulty may be distinguished from ungrammaticality (Chomsky (1965)). The distinction between knowledge of language and its use is commonly referred to as the distinction between *competence* and *performance*. In *Aspects of the Theory of Syntax*, Chomsky characterizes linguistic competence as:

> ...concerned primarily with an ideal speaker-listener, in a completely homogeneous speech community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance. [Chomsky (1965), p. 3]

The set of conditions that make up the restrictions on actual language use includes the list given in the above quotation, *i.e.*, "memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying knowledge of the language." Note that many of the factors involved in linguistic performance are independent of linguistic processing. For example, shifts of attention can be better explained by psychological theories of other components of the mind. Each of the independent conditions necessitates its own theory. These theories may then interact with the theory of sentence processing to give a complete theory of linguistic performance.[1] Sentences which violate a principle or rule of

[1]These theories will also interact with each other to give theories of other cognitive processes. A theory of attention, for example, would have effects on a theory of visual processing as well as other theories of cognitive abilities, including a theory of language processing.

grammar are called *ungrammatical*, while processing difficulty is usually measured in terms of *acceptability*. That is, a sentence which cannot be processed independent of its grammatical status is referred to as *unacceptable*.

### 1.1. Empirical Issues: Processing Breakdown

Most sentences that people encounter in everyday life are not difficult to process. However, there are a number of circumstances in which processing becomes difficult. A psychologically plausible model of linguistic processing must be able to account for not only the ease with which people understand most sentences but also the difficulty that people have with others. The processing breakdown effects to be accounted for by the theory presented in this thesis consist of *garden-path* effects and *processing overload* effects. A sentence which causes a garden-path effect is a grammatical sentence that has a local ambiguity in which there is a preference for one reading, but that reading turns out to be incompatible with the rest of the sentence, so that reanalysis is necessary to get a successful interpretation of the sentence.[2] I will concentrate here on those sentences for which the reanalysis reaches a conscious level: those sentences which people are aware of having difficulty with. As a result, when I use the term *garden-path effect* from this point forward, I will be referring to an effect associated with conscious difficulty (Bever (1970)). Evidence for the existence of garden-path effects comes from both psycholinguistic experiments and intuitions. For example, the processing of (3) (from Marcus (1980)) results in a garden-path effect:[3]

(3) # The cotton clothing is made of grows in Mississippi.

Sentence (3) is locally ambiguous at the point of processing the noun *clothing*: this noun may attach as head of a single noun phrase *the cotton clothing* or it may attach as subject of a relative clause that modifies the NP *the cotton*. People strongly prefer the single NP attachment at this point in the parse. Thus a garden-path effect results since it is the relative clause reading which is necessary to obtain a grammatical parse of the sentence. If the local ambiguity is not present, no comparable parsing difficulty ensues, as is demonstrated by (4):

(4) The cotton which clothing is made of grows in Mississippi.

Since the *wh*-word *which* separates the words *cotton* and *clothing*, these two words cannot unite to form a single noun phrase and the parsing difficulty in (3) is avoided.

While a psychologically plausible model must account for the conscious difficulty that people experience with many garden-path sentences, it also must account for the ease with which people process other sentences which are also locally or even globally ambiguous. For

---

[2]Related to garden-path effects are preferred readings for globally ambiguous sentences. If both the locally preferred and unpreferred readings of an ambiguity are compatible with the rest of the input string, then a preferred reading for an ambiguous sentence results.

[3]I will prefix sentences that cause conscious difficulty with the symbol "#". Hence garden-path sentences will be prefixed with the symbol "#", as will be sentences that are difficult to process independent of whether they contain ambiguity.

example, consider the pun in (5) (from Norvig (1988)):

(5) She criticized his apartment, so he knocked her flat.

The fact that (5) is amusing is due in part to the fact that *her flat* is ambiguous between two interpretations: 1) an adjectival predicate interpretation in which the NP *her* receives a thematic role from the adjectival reading of *flat*; and 2) a noun phrase interpretation in which *her* is a genitive pronoun modifying the nominal reading of *flat*.[4] A psychologically plausible model of sentence processing must account for the ease with which people recognize both possible meanings in (5).

Similarly, the sentences in (6) are locally ambiguous until the fourth word in each, but neither completion gives the human processor difficulty (Frazier & Rayner (1987)):

(6) a. The desert trains are especially tough on young people.
b. The desert trains young people to be especially tough.

There are two interpretations of the words *the desert trains*: a noun phrase reading and an incomplete sentential reading. Neither continuation causes any noticeable problems for the human processor, and this fact must be accounted for by a psychologically plausible model of parsing.

The second major source of processing breakdown data will be referred to as *processing overload* effects. A sentence that falls into this category is a grammatical sentence that is difficult to interpret, independent of how many local ambiguities it contains. Thus even if there is only one possible reading for a sentence in this class, it is still very difficult to obtain this interpretation. Sentences that induce processing overload effects therefore differ from sentences that induce garden-path effects since the difficulty in a garden-path sentence derives from a local ambiguity. Once an individual knows which local interpretation to follow in a garden-path sentence, the sentence becomes easy to process. The same is not true of a sentence inducing a processing overload effect: even if the individual knows what local interpretations to follow, the sentence is still difficult to understand. An example of a sentence which induces a processing overload effect is the following doubly center-embedded structure:[5]

(7) # The man that the woman that the dog bit saw likes fish.

Despite the fact that there is no local ambiguity in (7), it is still very difficult to understand. Furthermore, under conventional assumptions, (7) is not ungrammatical, as can be ascertained when it is compared with (8):

---

[4]The humor in (5) crucially relies on the presence of a nominal lexical entry for *flat* which is roughly synonymous with *apartment*. Most British people and some North Americans have this lexical entry.

[5]A tree structure is said to be *center-embedded* if a node which is on neither the left or right perimeter of the tree has further left and right subcomponents. A structure is said to be *multiply center-embedded* if it contains a center-embedded structure nested inside another center-embedded structure.

(8) The man that the woman likes eats fish.

The structure for (7) is made up of exactly the same kinds of components as is the structure for (8), only in more quantity. Since sentence (7) is grammatical, no principle of the grammar can rule out (8) (Chomsky (1956), Chomsky & Miller (1963), Chomsky (1965)). Thus (7) must be grammatical. In Chomsky's (1965) terms, (7) is grammatical, yet unacceptable.

It was noted earlier that a theory of human sentence processing may be useful to computer scientists for improving parsing efficiency. A theory of sentence processing which correctly predicts which linguistic inputs will give people difficulty has a number of additional natural language processing applications. Natural language generation (*e.g.*, in machine translation) is one such area in which processing breakdown effects should be taken into account. It would be undesirable to generate linguistic output for human consumption which people find difficult to process. For example, it would be undesirable to generate sentence (3), since it induces a garden-path effect. Sentence (4) would be much more appropriate:

(3) # The cotton clothing is made of grows in Mississippi.
(4) The cotton which clothing is made of grows in Mississippi.

Similarly, it would be undesirable for an automatic natural language generator to produce a multiply center-embedded sentence like (7). Rather, the information in (7) should be output in such a way that the center-embedding is limited. One possible acceptable sentence containing the same information as (7) uses the passive voice to avoid center-embedding:

(9) The man that was seen by the woman that the dog bit likes fish.

A further application of a theory of human linguistic processing breakdown is in the area of programs to check the linguistic structure of human-generated documents. For the same reasons that it is undesirable for a computer program to generate sentences like (7) or (3) it is undesirable for people to have sentences like these in their written documents. Clearly a theory of linguistic processing which predicts which sentences cause processing difficulty is of great use in such an application.

### 1.2. Sentence Processing with Limited Resources

The fact that certain constructions are difficult for the human linguistic processor should come as no surprise. The human parser has to process language in real time with limited resources: cutting corners is necessary to cope with the difficulty of the problem. In order to avoid the demand on resources, I hypothesize that the human parser heuristically determines its options based upon the evaluation of the possible linguistic representations of the input string with respect to a number of predetermined properties. Each of these properties is assumed to be associated with a certain weight or cost measured in terms of an abstract unit: the *processing load unit* (PLU). The total cost associated with a linguistic structure is assumed to be determined by adding together the weights associated with all the properties of that structure. These properties are defined such that the higher the weight of a structure, the less preferred that structure is.

Given the fact that the human parser has access to a limited quantity of resources, I hypothesize that there exists a maximal processing load (in PLUs) that the human parser can maintain. In other words, it is hypothesized that there exists a constant, $K$, such that the human linguistic processor maintains only those structures whose processing loads are less than or equal to $K$ PLUs. Thus a structure becomes unacceptable at a particular parse state if the combination of the processing costs associated with its lexical, syntactic, semantic and pragmatic properties at that state corresponds to a load greater than the processing overload constant $K$, in PLUs:[6]

(10)

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

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 that are associated with processing load,
$A_i$ is the number of times property $i$ appears in the structure.

Sentences which cause processing overload effects such as (7) can be explained by the existence of the overload constant $K$. Following Chomsky (1956, 1959a, 1959b, 1965), Bar-Hillel (1961), Chomsky & Miller (1963) and Langendoen (1975), I assume that the unacceptability of sentences like (7) is caused by short term memory constraints on the human parser. Under the assumptions given here, there exists a state in the parse of (7) such that the combination of the properties of the structure(s) at that state correspond to a processing load greater than the maximum allowable load, $K$ PLUs.

Similarly, given a parse state of an acceptable sentence, the combination of the loads associated with the properties of the structures at that state produces a load which does not exceed the maximum allowable processing load. Thus there will be a system of inequalities associated with each acceptable sentence, where each inequality has the form of (10), with the predicate "$\leq$" substituted for the predicate "$>$".

[6]In previous work, I have hypothesized that processing load units as described here correspond directly to units of linguistic short term memory (Gibson (1990a, 1990b)). I make no such claim here. Rather, it is assumed only that PLUs are *motivated* by short term memory, in that they give the parser a metric upon which to evaluate linguistic structures, so that some can be preferred over others. While it may turn out that there is a more direct link between processing load units and short term memory units, there currently exists no solid evidence in favor of such a hypothesis. As a result, I will assume that PLUs are motivated by short term memory, but do not necessarily correspond to short term memory units. See, however, King & Just (submitted) and MacDonald, Just & Carpenter (submitted) for evidence that suggests that people with larger short term memory capacities can process complex sentences more easily than people with smaller short term memory capacities. King & Just demonstrate such an effect on unambiguous relative clause constructions, while MacDonald, Just & Carpenter demonstrate a related effect on some garden-path constructions.

(11)

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

Furthermore, the assumption that each linguistic structure is associated with a processing cost provides a framework for the explanation of garden-path effects and preferred readings for ambiguous sentences. It is assumed that the language processor is an automatic device that uses a greedy algorithm: only the best of the set of all compatible representations for an input string are locally maintained from word to word. One way to make this idea explicit is to assume that restrictions on memory allow at most one representation for an input string at any time (see, for example, Frazier (1979), Frazier & Fodor (1978), Frazier & Rayner (1982), Rayner, Carlson & Frazier (1983), Ferreira & Clifton (1986), Pritchett (1988) and Abney (1989), *cf.* Marcus (1980) and Berwick & Weinberg (1984)). This hypothesis, known as the serial hypothesis, is easily compatible with the above view of processing load calculation: given a choice between two different representations for the same input string, simply choose the representation that is associated with the lower processing load.

However, the serial hypothesis is just one way of placing local memory restrictions on the parsing model. In this thesis I present an alternative formulation of local memory restrictions within a parallel framework. The model that I propose is essentially an elaboration of the *ranked parallel* approaches of Kurtzman (1985) and Gorrell (1987).[7]

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. For arguments and evidence in favor of the parallel hypothesis, see Crain & Steedman (1985), Kurtzman (1985), Gorrell (1987), Schubert (1984, 1986), Waltz & Pollack (1985), Norvig (1988), Altmann (1988), Altmann & Steedman (1988), Taraban & McClelland (1988), Ni & Crain (1989) and Trueswell *et al* (1989). For a history of the serial/parallel debate see Garnham (1985), Kurtzman (1985) or Gorrell (1987). Data that is normally taken to be support for the serial hypothesis includes 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 can also be accounted for in a parallel framework.

As a result of the plausibility of a parallel model, I propose to limit the *difference in processing load* that may be present between two structures for the same input, rather than limit the *number* of structures allowed in the processing of an input (*cf.* Gibson (1987), Gibson & Clark (1987), Clark & Gibson (1988)). Thus I assume that the human parser prefers one structure over another when the relative processing load (in PLUs) associated with maintaining the first is markedly lower than the processing load associated with the second. That is, I assume there exists some arithmetic preference quantity $P$ corresponding to a processing load, such that if the processing loads associated with two representations

[7]See also Schubert (1984, 1986), Waltz & Pollack (1985), Norvig (1988) and Hobbs *et al* (1988) for similar approaches. Crain & Steedman (1985) also assume a parallel architecture like that assumed here, although my assumptions regarding syntactic effects are quite different from theirs. See Section 2.6.

for the same string differ by relative load $P$, then only the representation associated with the smaller of the two loads is pursued.[8]

Given the existence of a preference factor $P$, garden-path effects can now be explained once the relationship between reanalysis and conscious processing difficulty is determined. Pretheoretically, there exist many possible relationships between reanalysis and conscious difficulty. The hypothesis that I will assume here is motivated by the fact that the primary component of parsing is the formation of larger structures from smaller ones: attachment. Thus it is reasonable to assume that expensive (conscious) reanalysis of an input string consists of reanalysis that takes place after some attachment has succeeded. In other words, reanalysis of previously attached input is expensive. For example, it may be that a word can be analyzed as two distinct categories, the difference in whose processing loads is high enough to cause a local preference for one over the other. Thus only the cheaper analysis will be initially pursued. However, if this analysis cannot take part in any attachments, it is reasonable to assume that reanalysis will not be costly here.[9] This hypothesis is given as follows, the Conscious Reanalysis Hypothesis:

(12) Conscious Reanalysis Hypothesis:
Reanalysis of an input string is expensive if and only if representations for some segment of that string have been attached in some previous parse state.

Given the existence of a preference factor $P$ together with a relationship between reanalysis and conscious processing difficulty, garden-path effects and preferred readings for ambiguous sentences can now be explained.[10] Both effects occur because of a local ambiguity which is resolved in favor of one reading. In the case of a garden-path effect, the locally favored reading is not compatible with the whole sentence, so that expensive reanalysis is necessary to arrive at an interpretation for the sentence. Given two representations for the same input string that differ in processing load by at least the factor $P$ PLUs, only the less computationally expensive structure will be pursued. 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. Thus if we know where one reading of a (temporarily) ambiguous sentence becomes the strongly preferred reading, we can write an inequality associated

---

[8] It is possible that the preference factor is a *geometric* one rather than an *arithmetic* one. Given a geometric preference factor, one structure is preferred over another when the ratio of their processing loads reaches that threshold value. I explore only the arithmetic possibility here; it is possible that the geometric alternative gives results that are as good, although I leave this issue for future research. See Chapter 10 for further comments on this and related issues.

[9] The word *that* is one example of a ambiguous word whose representations differ by more than $P$ PLUs under the assumptions and properties given in this thesis. See Section 6.9.5 for a description of the relevant data).

[10] It might also be possible to assume the serial hypothesis in conjunction with the preference constant $P$ proposed here: conscious difficulty occurs only when the difference between processing loads gets as high as $P$ PLUs; all other reanalysis is unconscious (inexpensive). While this hypothesis is possible, I will not pursue it here: I leave this investigation to future research.

with this preference:

(13)
$$\sum_{i=1}^{n} A_i x_i - \sum_{i=1}^{n} B_i x_i > P$$

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 that are associated with processing load,
$A_i$ 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.

Similarly, if more than one structure is maintained in a particular parse state, then the difference between these processing loads must be less than or equal to the load difference quantity $P$ PLUs:[11]

(14)
$$\sum_{i=1}^{n} A_i x_i - \sum_{i=1}^{n} B_i x_i \leq P$$

The assumptions and methodology described here also provide simple explanations of gradations in acceptability as well as plausible partial explanations of gradations in garden-path effects and ease of recognition of ambiguity. For example, one sentence is harder to process (less acceptable) than another if its processing load at some parse state is greater than the processing loads at all parse states of the second sentence. Similarly, one factor that may contribute to the relative ranking of garden-path effects is the difference between processing load differences at the point of local ambiguity resolution: one garden-path sentence is more difficult to parse than another if the processing load difference at the point of local ambiguity resolution is greater than that associated with the second.[12]

Given a parsing algorithm together with $n$ properties and their associated processing loads $x_1 ... x_n$, we may write inequalities having the form of (10), (11), (13) and (14) cor-

[11]While the load difference between two structures may be less than $P$ PLUs so that both are maintained, it may still be the case that one of these structures is processed faster than the other, without a conscious effect. For example, experiments performed by Frazier & Rayner (1982), Ferreira & Clifton (1986) and Trueswell, Tanenhaus & Garnsey (submitted) among others indicate that some analyses for the same input string take longer to arrive at than others, despite a lack of conscious effect. These so-called "unconscious garden-path effects" can be explained in a parallel framework by appeal to access time of more heavily weighted structures: the higher the weight of a structure, the slower it is to access (*cf.* Kurtzman (1985)). Thus one structure may be processed more slowly than another without the effect reaching a conscious level.

[12] The level of difficulty in parsing a garden-path sentence is probably also due to some function of the number of words over which backtracking must take place as well as what structures are common to both the locally preferred and unpreferred readings. Other factors, such as context, are also probably involved. In any case, the relative difficulty of parsing garden-path sentences is very likely explained by more than just difference in processing load (Frazier & Rayner (1982)).

responding to the state of processing load at various parse states. An algebraic technique called *linear programming* can then be used to solve this system of linear inequalities, giving an $n$-dimensional space for the values of $x_i$ as a solution, any point of which satisfies all the inequalities.

In this thesis I will concentrate on syntactic properties.[13] In particular, I present four properties. The first three properties – the Property of Thematic Reception (PTR), the Property of Lexical Requirement (PLR) and the Property of Thematic Transmission (PTT) – are based on principles from Government and Binding Theory (Chomsky (1981, 1986a)): the $\theta$-Criterion and the Projection Principle respectively.[14] The fourth – the Property of Recency Preference (PRP) – prefers structures resulting from local attachments over those that result from more distant attachments (*cf.* Kimball (1973), Frazier (1979)). It will be shown that these properties, once associated with processing loads, predict a large array of garden-path effects. Furthermore, it is demonstrated that these properties also make desirable predictions with respect to unacceptability due to memory capacity overload.

The organization of the thesis is given as follows. The second and third chapters describe previous research relevant to the issues undertaken here: Chapter 2 gives an overview of previous research in the processing of ambiguity resolution; Chapter 3 describes previously presented explanations of processing overload phenomena. Chapter 4 describes the structure of the underlying parser to be assumed here. In Chapter 5 the Properties of Thematic Reception, Lexical Requirement and Thematic Transmission are proposed. Chapter 6 examines a large number of locally ambiguous sentences with respect to these properties, and a solution space for the processing loads of the properties is calculated in terms of the Preference Factor $P$. In Chapter 7 the Property of Recency Preference is presented. Further locally ambiguous sentences are examined with respect to this property and the Properties of Thematic Reception and Lexical Requirement: a solution space for values of the processing loads of these properties in terms of the Preference Factor $P$ results. In Chapter 8 it is shown that the resultant space seems to make the right predictions with respect to processing overload effects. Chapter 9 gives some data that syntactic approaches to preference phenomena do not handle easily, as well as suggestions for possible non-syntactic load-bearing properties to explain this data. Concluding remarks and suggestions for further research can be found in Chapter 10.

---

[13] Note that I assume that there also exist lexical, semantic, pragmatic and discourse-level properties which are associated with significant processing loads.

[14] In fact, any syntactic theory which contains correlates of the relevant parts of the $\theta$-Criterion and Projection Principle will also derive the Properties of Thematic Reception, Lexical Requirement and Thematic Transmission. See footnote 80 in Chapter 5 for more on this issue.

## 2. Previous Work: Ambiguity Resolution

### 2.1. Bever's Heuristic Strategies

One of the first to give a theory of parsing preferences was Bever (1970) (see also Fodor, Bever & Garrett (1974)). In his work he proposed that the human parser operates using a number of heuristics, one of which, the *Canonical Sentoid Strategy*,[15] is given in (15):

(15) The Canonical Sentoid Strategy: The first N...V...(N)... clause is the main clause, unless the verb is marked as subordinate (Bever (1970) p. 294).

This heuristic accounts for the fact that people experience garden-path effects in the following well-known examples (both originally from Bever (1970)):

(16)a. # The horse raced past the barn fell.
b. # The editor authors the newspaper hired liked laughed.

The heuristic strategy in (15) predicts that people analyze the words *the horse raced* in (16a) as the main clause in the sentence. When it turns out that this analysis is incorrect, reanalysis is necessary and a garden-path effect results. Similarly, the strategy in (15) predicts that the words *the editor authors the newspaper* in (16b) will be initially analyzed as a main clause. This preliminary analysis also results in a failed parse, and a garden-path effect is predicted for (16b) as desired.

While Bever's strategies make correct predictions for some sentences, much data is also left unaccounted for. For example, consider (17):

(17) # The patient persuaded the doctor that he was having trouble with to leave.

People locally prefer to attach the complement clause *that he was having trouble ...* in (17) as argument of the verb *persuade* rather than as modifier of the NP *the doctor*. A garden-path effect results, since it is the NP modifier attachment that is necessary for a grammatical parse of the sentence. However, none of Bever's heuristics accounts for this preference.

The heuristics also fail to predict the preferred reading in (18):

(18) John put the book on the table in the hall.

The preferred attachment of the prepositional phrase *on the table* is as argument to the verb *put* rather than as a modifier of the NP *the book*. None of Bever's strategies applies to cases like (18), and consequently no prediction is made.

There exist many further cases of garden-path effects or preferred readings that are not

[15] This strategy is called simply "Strategy B" in Bever (1970). The name "Canonical Sentoid Strategy" first appears in Fodor, Bever & Garrett (1974).

predicted by Bever's strategies. Furthermore, the heuristics are sometimes incorrect in predicting difficulty. Consider the sentences in (19) (from Frazier & Rayner (1987)):

(19)a. The desert trains are especially tough on young people.

b. The desert trains young people to be especially tough.

The words *the desert trains* in each of the sentences in (19) are ambiguous between noun phrase and sentential readings. Bever's strategy in (15) predicts that people should prefer the sentential reading, and thus have difficulty with (19a). However, no such difficulty occurs.

Many further such examples of both garden-path over-prediction and under-prediction of Bever's strategies exist, so that this theory is no longer considered a satisfactory explanation of parsing difficulty. However, this work was one of the first testable theories of human linguistic performance, a great step beyond all previous work.

## 2.2. The Principles of Minimal Attachment and Late Closure

Taking principles proposed by Kimball (1973) as a starting point, Frazier (1979) developed the principles of *Minimal Attachment* and *Late Closure*, which make up perhaps the best-known theory of preferred readings and garden-path effects today (see also Frazier & Fodor (1978), Fodor & Frazier (1980), Frazier & Rayner (1982), Rayner, Carlson & Frazier (1983), and Frazier (1990)).[16,17] This theory assumes the serial hypothesis: that at most one representation can be maintained for the input string at each parse state. In order to decide which structure to choose at a given parse state, the parser invokes the principles of Minimal Attachment and Late Closure. The principle of Minimal Attachment is given as follows (Frazier (1979), Frazier & Rayner (1982)):

(20) Minimal Attachment: Attach incoming material into the phrase-marker being constructed using the fewest nodes consistent with the well-formedness rules of the language.

Frazier assumes a phrase structure related to $\overline{\text{X}}$ Theory (Jackendoff (1977), Chomsky (1986b)) in which arguments are attached as sisters to their heads, specifiers are attached as sisters to one-bar levels and modifiers are adjoined to one and two-bar levels (see Section 4.3.1). Crucially, modification involves adjunction and therefore the creation of extra nodes, while argument and specifier attachments do not. Thus, by the principle of Minimal Attachment, argument and specifier attachments are always preferred over modifier

[16] The principles from Kimball (1973) that Frazier based her work on are called the Principles of New Nodes and Right Association. I will not be discuss these principles here since they are rather similar to their successors, Minimal Attachment and Late Closure. For discussion of Kimball's original principles, see Kimball (1973, 1975), Frazier & Fodor (1978), Frazier (1979) and Church (1980).

[17] See Wanner (1980, 1987) for a description of scheduling principles inside an Augmented Transition Network parser that give results like those of Minimal Attachment and Late Closure. Also see Pereira (1985) and Shieber (1983) for descriptions of similar scheduling principles inside a shift-reduce parser.

attachments.[18]

The principle of Late Closure is given below (Frazier (1979), Frazier & Rayner (1982)):

(21) Late Closure: When possible, attach incoming lexical items into the clause or phrase currently being processed (*i.e.*, the lowest possible nonterminal node dominating the last item analyzed).

In addition, it is assumed that the principle of Minimal Attachment applies before the principle of Late Closure, so that Minimal Attachment has priority when conflicts between the two arise.[19] The principles of Minimal Attachment and Late Closure inside a serial processing model correctly predict a large array of garden-path effects and preferred readings of ambiguous input. Consider the following examples with respect to Minimal Attachment:

(22)a. # The horse raced past the barn fell.
b. # The patient persuaded the doctor that he was having trouble with to leave.

Minimal attachment predicts the garden-path effects in the above sentences. Consider the parse of (22a) at the point of parsing the word *raced*:

(23)a. $[_S\ [_{NP}$ the horse $]\ [_{VP}$ raced $]]$
b. $[_S\ [_{NP}$ the $[_{N'}\ [_{N'}$ horse$_i$ $]\ [_{S'}\ [_S\ [_{VP}$ raced $e$ $]]]\ ]]\ ]$

Many fewer nodes need to be constructed for the matrix verb reading, (23a), than for the reduced relative clause reading, (23b), so the matrix verb reading is preferred. Since it is the reduced relative reading that is necessary for a grammatical parse of (22a), a garden-path effect results.

Similarly, at the point of parsing the complementizer *that* in (17) fewer nodes need to be constructed for the argument attachment than for the NP modifier attachment:

---

[18] Frazier's assumptions regarding phrase structure are not, in fact, clear. The use of any grammar which is not linguistically motivated requires further explanation, and is thus stipulative. Frazier's initial grammatical assumptions suffered from the difficulty of being stipulative in that she assumed nominal adjuncts were adjoined while verbal adjuncts were not. Church (1980), Schubert (1984), Hirst (1987) and Abney (1989) among others note this problem with her theory. In particular, Abney (1989) points out the stipulativeness of her phrase structure assumptions and gives an alternative that makes explicit the preference for argument (thematic) attachments over non-thematic attachments (see also Gibson (1987), Clark & Gibson (1988) for similar proposals). Frazier (1990) responds to Abney's criticisms by stating explicitly that thematic attachments do not involve the creation of extra nodes, implying that modifier attachments do, and hence adopting an $\overline{\text{X}}$-like schema. Thus, in effect, she takes these criticisms as valid.

[19] In the description of the Sausage Machine (Frazier & Fodor (1978)), Late Closure effects are assumed to be derivable from limiting the viewing region of the parser to five or six words. However, Wanner (1980) gives evidence that demonstrates that limiting the parser's viewing window would not get all the desired effects. In order to account for Wanner's data, Fodor & Frazier (1980) hypothesize that Local Association (Late Closure) is of full principle status, thus making the viewing window hypothesis somewhat superfluous. See, however, Fodor & Frazier (1980) for arguments that the viewing window is still necessary.

(24)a. $[_S\ [_{NP}$ the patient $]\ [_{VP}\ [_{V'}\ [_V$ persuaded $]\ [_{NP}$ the doctor $]\ [_{S'}$ that $]\ ]]\ ]]]$
b. $[_S\ [_{NP}$ the patient $]\ [_{VP}\ [_{V'}\ [_V$ persuaded $]\ [_{NP}$ the $[_{N'}\ [_{N'}$ doctor$_i$ $]\ [_{S'}$ that $[_S\ ]\ ]]]\ ]]]$

The argument attachment, (24a), is preferred and a garden-path effect results since it is the NP modifier attachment that is necessary in order to arrive at a grammatical structure for (17).

The principle of Late Closure predicts the garden-path effect in (27a) as well as the preferred reading in (27b) (from Wanner (1980)):[20]

(27)a. # Since she jogs a mile seems light work.
b. Bill thought John died yesterday.

In (27a), there is a syntactic ambiguity at the point of parsing the noun phrase *a mile*: this NP may attach as either direct object of the verb *jogs* or as subject of the matrix clause to follow:

(28)a. $[_S\ [_{S'}$ since $[_S\ [_{NP}$ she $]\ [_{VP}\ [_{V'}\ [_V$ jogs $]\ [_{NP}$ a mile $]\ ]]]]\ [_S\ ]]$
b. $[_S\ [_{S'}$ since $[_S\ [_{NP}$ she $]\ [_{VP}$ jogs $]]]\ [_S\ [_{NP}$ a mile $]\ ]]$

Each of these attachments involves the same number of new nodes, so the principle of Minimal Attachment does not apply here. However, since the direct object attachment is within the same clause as the preceding words, this attachment is preferred by the principle of Late Closure. However, the matrix subject attachment is the attachment that is necessary

[20] Note that if (27a) contained a comma after the word *jogs*, there would be no question of grammaticality:

(25) Since she jogs, a mile seems light work.

Similarly, an intonation break would effect disambiguation in the spoken language equivalent of (25). However, without the comma or intonation break, there is some question in the literature as to whether sentences like (27a) are ungrammatical or merely difficult to process (see, for example, Frazier & Rayner (1982) and Kurtzman (1985) on this issue). Following Frazier & Rayner (1982), I argue that these sentences are in fact grammatical. Suppose that sentences of this type were not grammatical, so that some principle of the grammar rules them out. If such a principle existed, it would need to rule out each sentence which 1) contains a pre-matrix sentential modifier and 2) lacks a comma separating the pre-matrix modifier from the matrix clause. It would not be possible for the grammar to force a comma only if local ambiguity existed without the comma, since principles of grammar do not have access to knowledge of ambiguity. The detection of ambiguity is a performance issue, not a competence one. Hence the principle of grammar that rules out (27a) would also rule out the sentences in (26):

(26) a. Ever since Mary met John she has been eating only pizza.
b. While Fred was talking to Sue he was thinking of the algebra quiz.

First of all, it should be noted that these are not contrived examples. Sentences like the ones in (26) occur frequently in written text and are almost universally judged to be both grammatical and acceptable. Note that there is no possible ambiguity when the matrix clause subject of each of these sentences appears. While placing a comma between the presentential modifier and the matrix subject is permitted in these sentences, it is not obligatory. Since the sentences in (26) are grammatical, (27a) must also be grammatical. That is, any principle of grammar ruling out (27a) for its lack of punctuation would also rule out the sentences in (26), an undesirable consequence. Hence the difficulty in processing (27a) must therefore be due to performance: a garden-path effect. See also footnote 125.

for a successful parse of (27a). Thus a garden-path effect is correctly predicted.

Similarly, the nominal adverb *yesterday* can attach to either the embedded clause or to the matrix clause in (27b):

(29)a. $[_{IP}$ Bill $[_{VP}$ thought $[_{IP}$ John $[_{VP}$ $[_{VP}$ died $]$ $[_{AdvP}$ yesterday $]$ $]]]]$

b. $[_{IP}$ Bill $[_{VP}$ $[_{VP}$ thought $[_{IP}$ John $[_{VP}$ died $]]]$ $[_{AdvP}$ yesterday $]$ $]]$

As in (28a), each of these attachments involves the same number of new nodes, so the principle of Minimal Attachment does not apply. Since the embedded clause occurs more recently in the input string, this attachment is preferred. Thus there is a preferred reading of (27b): that which associates *yesterday* with the embedded clause.

Minimal Attachment necessarily applies before Late Closure in order to account for the preferred readings of sentences like (30):

(30) John put the book on the table in the hall.

In this sentence, the preferred attachment site for the prepositional phrase *on the table* is as argument of the verb *put*, as Minimal Attachment predicts. However, the principle of Late Closure by itself predicts that the preferred attachment site of this PP is as modifier of the NP *the book*, since the NP *the book* is the most recent available attachment site in the input. This prediction contradicts the prediction of Minimal Attachment. Since Minimal Attachment is in fact correct in this and other similar situations, it is assumed that Minimal Attachment is applied first, thus taking precedence when the predictions of the two principles conflict. While this ranking of principles gets the desired effects, it is theoretically undesirable, since such a ranking is not adequately independently motivated. Thus the theory would be improved if the two principles did not need to ordered with respect to each other.[21]

### 2.2.1. Problems with Minimal Attachment and Late Closure

While the principles of Minimal Attachment and Late Closure correctly predict many garden-path effects and preferred readings, there are a number of difficulties with this theory.[22]

First of all, the theory of Minimal Attachment did not originally depend on thematic relations: it depended on how the phrase structure grammar for the language was written (Church (1980), Schubert (1984), Hirst (1987), Abney (1989), McRoy & Hirst (1990)). However, no explicit theory of writing grammar rules was given. Thus rules could be written in such a way so that Minimal Attachment had the right effects, resulting in a stipulative

[21]Indeed, part of the purpose of the Sausage Machine (Frazier & Fodor (1978)) was to eliminate the ranking of Kimball's (1973) processing principles. See footnote 19.

[22]Note that the empirical difficulties with Minimal Attachment and Late Closure that are observed in this section also apply to Kimball's original principles as well as the ATN and shift-reduce parser scheduling principles discussed in Wanner (1980), Pereira (1985) and Shieber (1983).

theory. For example, consider adjunct prepositional phrase attachment. It happens that locative PPs usually prefer to attach to phrases that represent events over those that do not. Consider, for example, (31):

(31) Mary read the article in the {bathtub, magazine}.

People usually have less difficulty processing the noun *bathtub* rather than the noun *magazine* in (31) indicating that VP modification of the PP headed by *in* is preferred. Since events are usually lexically realized as verbs and not as nouns, it is possible to write a set of phrase structure rules that will give the right effects under Minimal Attachment:

VP → V NP (PP*)
NP → N
NP → NP PP

By allowing non-argument prepositional phrases at the same level as argument phrases in the verb phrase rules, additional levels of VPs are avoided. Requiring that the same non-argument PPs adjoin to noun phrase constituents guarantees a greater number of nodes in an NP attachment given the option of a VP attachment. Thus given these rules under Minimal Attachment, the appropriate predictions are made with respect to sentences like (31). However, this theory has little predictive power if otherwise unmotivated rules are written to account for each effect. No current grammatical theory distinguishes verbal modifiers from nominal modifiers as in the above grammar since there is no linguistic motivation for making such a distinction. Thus the grammars behind Minimal Attachment have the potential to be stipulative.

However, even if Minimal Attachment is based on a linguistically sound theory of grammar (*e.g.*, lexically determined thematic role assignment together with $\overline{\text{X}}$ Theory as is suggested by Frazier (1990)), this theory of parsing preferences still has numerous empirical shortcomings. First, consider the sentences in (32) ((32b) and (32c) from Frazier & Rayner (1982)):

(32)a. Bill knew John liked Mary.
b. The city council argued the mayor's position was incorrect.
c. Tom heard the gossip wasn't true.

Each of the matrix verbs in the sentences in (32) is ambiguous between a reading which subcategorizes for a simple noun phrase and one which subcategorizes for a sentential complement. Thus each of these sentences is ambiguous at the point of processing the noun phrase immediately following the matrix verb. Consider structures for the relevant ambiguity in (32a):

(33)a. $[_S\ [_{NP}$ Bill $]\ [_{VP}$ knew $[_{NP}$ Mary $]]]$
b. $[_S\ [_{NP}$ Bill $]\ [_{VP}$ knew $[_{S'}\ [_S\ [_{NP}$ Mary $]\ ]]]]$

In (33a), the NP *Mary* attaches directly as object of one reading of the verb *knew*. In (33b), the NP *Mary* attaches as subject of the sentential complement of another reading of the verb *knew*. Since structure (33b) requires at least two more nodes than does structure (33a), the principle of Minimal Attachment dictates that only structure (33a) is maintained. As a result of this local preference, garden-path effects should result in all of the sentences in (32), since backtracking will be necessary in each to obtain the necessary readings.

In order to test this prediction, Frazier & Rayner (1982) had subjects read sentences like those in (32), along with a number of control sentences, while their eye movements were followed using an eye-tracking machine. Frazier and Rayner found that when the ambiguous region was large enough, the pattern of eye movements suggested that subjects were backtracking when they got to the point of disambiguation in sentences like (32), thus providing support for the Minimal Attachment hypothesis. Frazier & Rayner concluded that, although backtracking doesn't always reach a conscious level in sentences like (32), it still takes place, and thus the serial model with backtracking best explains the data.

However, the results of Frazier & Rayner (1982) have recently been drawn into question by experiments by Holmes, Kennedy & Murray (1987) and Kennedy *et al* (1989). In these experiments subjects read sentences like those in (32) as well as similar sentences in which the disambiguating complementizer *that* appears, as in the sentences in (34):

(34)a. Bill knew that John liked Mary.
b. The city council argued that the mayor's position was incorrect.
c. Tom heard that the gossip wasn't true.

The same regressive eye movements were observed in reading these sentences as in reading those in (32). Thus these researchers attribute the increased complexity that subjects have with sentences like (32) to the increased complexity of the structure that needs to be constructed. As a result, it appears that more than one structure for the ambiguous regions are being maintained, contrary to the serial hypothesis and Minimal Attachment. These results confirm experimental work performed by Kurtzman (1985) and Gorrell (1987) who also found evidence that people maintain multiple representations for ambiguous regions following verbs with more than one subcategorization, such as those in (32).

However, it is still possible that Frazier and her colleagues would argue that the experiments performed by Holmes, Kennedy & Murray (1987) and Kennedy *et al* (1989) were not sensitive enough to measure the backtracking that takes place in sentences like those in (32).[23] If this were the case, then it becomes necessary to explain why some garden-path

[23]In fact, Rayner & Frazier (1987) make this argument in their criticism of Holmes, Kennedy & Murray (1987). Rayner and Frazier attribute the difference between their own results and those of Holmes, Kennedy and Murray to the different tasks: Rayner and Frazier studied eye-movements, while Holmes, Kennedy and Murray used a self-paced reading study. Rayner and Frazier claimed that self-paced reading was not a sensitive enough task to detect the backtracking in the test sentences. As support for their claim, they replicated Holmes, Kennedy and Murray's experiments using an eye-tracker and once again found evidence of backtracking. However, their experimental method was flawed since many of the sentences presented to subjects included line-breaks, a complicating factor. Kennedy *et al* (1989) noted this flaw and performed the same experiment using an eye-tracker on data that included no line-breaks. Their results replicated the results of Holmes, Kennedy & Murray (1987), once again contradicting the predictions of the serial hypothesis and Minimal Attachment. However, this issue is still not closed until Kennedy *et al*'s results are

effects cause conscious difficulty, while others do not. Frazier & Rayner (1982) hypothesize that reanalysis is unconscious up to a certain threshold quantity, while beyond that threshold, reanalysis becomes consciously noticeable. Consider once again (32a):

(32a) Bill knew John liked Mary.

Breakdown is predicted by the Minimal Attachment model at the point of parsing the second verb, since no NP is available to be the subject for this verb. Frazier & Rayner hypothesize that the human parser recognizes how to repair the failed parse in (32a) by looking back in the input string for an NP that can serve as the subject for this verb. An appropriate NP *John* is found, and the resulting S node is attached in the position previously occupied by the NP. Frazier & Rayner hypothesize that this repair is fast and inexpensive, so that it is unconscious. They also note that the same reanalysis procedure does not work on (35):

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

Reanalysis is necessary in (35) upon reading the verb *fell*, since no subject is available for this verb. Searching back in the input string yield no plausible candidates for the subject of this verb so that further reanalysis is necessary. Presumably there is enough additional reanalysis so that people are consciously aware of it.

While this theory of reanalysis is plausible as far as it goes, it is too vague to be tested as presented, since no units for measuring reanalysis difficulty are proposed nor is a concrete threshold value given. For example, the resulting theory should be able to account for the difference between (32a) and (27a):

(27a) # Since she jogs a mile seems light work.

The reanalysis needed in (27a) is very like that in (32a), since the NP immediately preceding the point of breakdown, the second verb, is reanalyzed as the subject of this verb. However, (27a) induces a conscious garden-path effect while (32a) does not. This difference must be accounted for by any theory of reanalysis inside a serial framework.

Similarly the difference between (35) and (36) must be explained by the reanalysis theory:

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

While (35) causes a conscious garden-path effect, (36) does not, despite its near identical structure and ambiguity (Pritchett (1988)). While it may be possible to account for these differences inside a serial framework with reanalysis,[24] the theory of reanalysis cost has not yet been adequately formalized in the framework proposed by Frazier and her colleagues.[25]

---

replicated by other researchers.

[24]Pritchett (1988) gives one possible metric of reanalysis computation (see Section 2.3).

[25]An additional possible analysis of the contrast between (36) and (35) inside the Minimal Attachment framework might appeal to the difference in distance over which reanalysis takes place in each. That is,

A further empirical difficulty with the theory of Minimal Attachment and Late Closure is demonstrated by the contrast between the examples in (39):

(39)a. # The patient persuaded the doctor that he was having trouble with to leave.
b. The patient persuaded the doctor that was having trouble to leave.

The explanation of the difficulty in (39a) is given by Minimal Attachment. Fewer nodes are needed in the construction of the argument attachment of the complementizer *that* than for the adjunct attachment, so the argument attachment is pursued. A garden-path effect results when it is the adjunct attachment that is necessary. However, this explanation also predicts that (39b) should also induce a garden-path effect, contrary to fact. That is, if the garden-path status of (39a) is due to Minimal Attachment, then (39b) should induce a similar effect since the same local preference takes place.[26]

Additional empirical difficulties with Minimal Attachment are presented by the following data (from Fodor (1985)):

---

(36) might not be difficult to process although misanalysis occurs at the point of processing the verb *found*, because the mistake in processing is noticed very soon after its occurrence: at the first word following the verb. In contrast, the point at which misanalysis is first noted in (35) is long after the incorrect analysis is initiated, when the main verb *fell* is input. The difference in the proposed reanalysis initiation between the two sentences derives from the fact that the verb *found* is obligatorily transitive, while the verb *raced* is optionally intransitive.

However, such an explanation of the lack of difficulty in (36) must also account for the following garden-path sentences, all of which have the same property:

(37) a. # John gave the boy rocks hit a dollar.
b. # I sent the child books fell on a present.
c. # While the children were playing football was on television.
d. # While my roommate was cleaning equipment arrived in the yard.
e. # While the chicken cooked refreshments were served in the living room.
f. # In the desert trains arrive less often.
g. # Beside the warehouse fires break out frequently.

Each of these sentences induces a strong garden-path effect, and yet each includes local ambiguity that is resolved exactly one word after it occurs. Furthermore, the following garden-path sentences include local ambiguities that are resolved two words later:

(38) a. # I convinced her children are noisy.
b. # The horse raced quickly fell down by the barn.
c. # The Russian women loved died very recently.

Thus ease or difficulty of reanalysis cannot be the result of the distance over which reanalysis takes place by itself: some additional explanation for the difficulty processing the sentences in (37) and (38) must also be given.

Moreover, note that an account of the difference between (36) and (35) which allows a limited quantity of look-ahead rather than a limited quantity of backtracking suffers from the same difficulties as does the limited backtracking hypothesis. That is, if the difference between (36) and (35) is to be explained in terms of the parser's ability to look ahead in the input string one word, then the fact that the sentences in (37) cause garden-path effects can no longer be explained.

[26]Note that the distance over which reanalysis needs to take place is only one word in the processing of (39b), in contrast to (39a) where the reanalysis distance is longer. But, as noted in footnote 25, ease or difficulty of reanalysis cannot by itself be the result of the distance over which reanalysis takes place. See this footnote for evidence against such a hypothesis.

(40)a. The report that the president sent to us helped us make the decision.

b. The report that the president sent the troops into combat depressed me.

The sentences in (40) are both ambiguous at the point of processing the phrase introduced by complementizer *that*. In (40b), the complementizer is attached as an argument of the NP *the fact* and in (40a) it is attached as a relative clause modifier of the same NP. Minimal Attachment predicts that the argument attachment should be preferred, resulting in backtracking in the processing of (40a). However, no garden-path effect occurs in (40a). Thus the reanalysis algorithm that is needed in the theory of Minimal Attachment must account for the lack of difficulty experienced processing sentences like (40a).

A further area of difficulty for the principles of Minimal Attachment and Late Closure is in the area of lexical ambiguity. Consider sentence (41) (from Frazier (1979)):

(41) # Without her contributions would be inadequate.

Frazier (1979) suggests that Late Closure can account for the garden-path effect in (41) in much the same way that it explains the garden-path effect in (27a). There are two possible attachment sites for the NP *contributions*: one as head of the object NP for the preposition *without*, and the other as subject of the matrix clause to come. These two attachments are given in (42):

(42)a. $[_S$ $[_{PP}$ without $[_{NP}$ $[_{Det}$ her ] contributions ]] $[_S$ ]]

b. $[_S$ $[_{PP}$ without $[_{NP}$ $[_{Pro}$ her ]] $[_S$ $[_{NP}$ contributions ]]]

Structure (42a) includes categorization of the word *her* as a genitive determiner, while (42b) categorizes *her* as an accusative pronoun. The two structures each contain the same number of nodes, so that Minimal Attachment does not apply. However, the principle of Late Closure prefers (42a) over (42b) since in (42a) the NP *contributions* attaches to the phrase currently being parsed, the PP headed by *without*, while in (42b), *contributions* attaches to a new phrase, the main clause. While this explanation seems reasonable, it fails to account for the following similar examples:

(43)a. # Without the computer companies would be less productive.

b. # Without light cameras are useless.

c. # Without cotton clothing would be less comfortable.

Consider (43b). The word *light* is ambiguous between adjectival and nominal readings, giving the two following possible structures for the input string *without light cameras*:

(44)a. $[_S$ $[_{PP}$ without $[_{NP}$ $[_{N'}$ $[_{Adj}$ light ] $[_{N'}$ cameras ]]]] $[_S$ ]]

b. $[_S$ $[_{PP}$ without $[_{NP}$ $[_{N'}$ light ]] $[_S$ $[_{NP}$ cameras ]]]

The word *light* is an adjectival modifier of the noun *cameras* in (44a), and therefore adjoins to a projection of its head. On the other hand, the word *light* in (44b) serves as the direct object of the preposition *without*, so no adjunction takes place. As a result, there are fewer nodes in structure (44b) than in structure (44a) and thus Minimal Attachment predicts that (44b) should be locally preferred.[27] Thus Minimal Attachment predicts that people will have no trouble processing (43b). On the contrary, however, people strongly prefer structure (44a), so that a garden-path effect results. Similarly, Minimal Attachment predicts no difficulty with any of the sentences in (43).[28] In fact, Minimal Attachment incorrectly predicts that the sentences in (45) induce garden-path effects:

(45)a. Without the computer companies we would be less productive.
b. Without light cameras people would take fewer photographs.
c. Without cotton clothing I would be much less comfortable.

Consider now the following set of examples in which local lexical ambiguities can be resolved in two ways, neither of which gives the human parser difficulty:

(46)a. I gave her earrings on her birthday.
b. I gave her earrings to Sally.
c. The president brought the committee folders for their notes.
d. The president brought the committee folders to the secretary.
e. We gave the department computers on their 20th anniversary.
f. We gave the department computers to a charity.

Consider (46a) and (46b) at the point of parsing the word *earrings*. Minimal Attachment says nothing about a preference in this situation since the two possible structures have the same number of nodes. However, Late Closure predicts that the single NP reading of *her earrings* is preferred, thus incorrectly predicting that (46a) induces a garden-path effect.[29] In contrast, Minimal Attachment predicts that the double object NP attachment is the preferred reading of the initial strings in each of sentences (46c) – (46f), thus incorrectly predicting that (46d) and (46f) induce garden-path effects.

Next consider the following examples adapted from Frazier & Rayner (1987):

---

[27] In fact, Late Closure makes the right prediction in cases like those in (43), thus contradicting Minimal Attachment. However, Minimal Attachment takes priority over Late Closure in order to account for other examples (see *e.g.*, (30)).

[28] The structure for the NP *the computer companies* in (43a) is not clear-cut: the word *computer* may be categorized as a noun, as part of a noun-noun compound *computer companies*, or it may be categorized as an adjective derived from a noun. However, in either case the structure representing *computer* is adjoined to a projection of the NP headed by *companies*, so the argument presented above remains the same. The same argument applies for the NP *cotton clothing* in (43c).

[29] It is of course possible that reanalysis is involved in the sentences in (46), but that this reanalysis does not reach the conscious level. However, no adequate formalization of such a hypothesis exists within the Minimal Attachment literature. See, however, Section 2.3 for one possible metric of reanalysis computation.

(47)a. The warehouse fires numerous employees each year.

b. The warehouse fires harm some employees each year.

c. The desert trains young people to be especially tough.

d. The desert trains are especially tough on young people.

e. The county buses are filled with grade school children from another county.

f. The county buses most of the grade school children to another county.

Frazier & Rayner (1987) showed that people experience no garden-path effects when parsing sentences like those in (47), despite the lexical ambiguity in each. For example, the first three words *the warehouse fires* in (47a) and (47b) are ambiguous between noun phrase and sentential readings, as shown in (48):

(48)a. $[_S\ [_{NP}\ [_{Det}$ the $]\ [_{N'}$ warehouse $]]\ [_{VP}$ fires $]]$

b. $[_S\ [_{NP}\ [_{Det}$ the $]\ [_{N'}\ [_{NP}$ warehouse $]\ [_{N'}$ fires $]]]]$

Minimal Attachment predicts that structure (48a) will be preferred since this structure contains fewer nodes than does structure (48b). The difference in node count comes from the fact that *warehouse* is a pre-nominal modifier in (48b), and thus must be adjoined to a projection of its head. Thus Minimal Attachment incorrectly predicts that (47b), (47d) and (47f) should all induce garden-path effects.

Thus Frazier & Rayner hypothesize that Minimal Attachment does not apply when a categorical ambiguity is encountered. Instead, the parser is assumed to *delay* until the ambiguity can be resolved. Note that it is undesirable from a theoretical point of view to have a different parsing algorithm when lexical ambiguity is encountered; a more parsimonious theory would not have such a distinction. Furthermore, a delay strategy has empirical difficulties as is demonstrated by the following examples (adapted from the examples in (37) in footnote 25):

(49)a. # John gave the boy rocks hit a dollar.

b. # While the children were counting trains raced by the window.

c. # In the desert trains arrive less often.

d. # Beside the warehouse fires break out frequently.

Consider (49c). The word *trains* is ambiguous between nominal and verbal readings, and thus the parser should delay at this point. The ambiguity is resolved in processing the next word, and thus a delay strategy predicts no difficulty with this sentence. However, this sentence induces a garden-path effect, contrary to prediction. The other examples in (49) are similarly predicted to be easy by the delay strategy, but all induce garden-path effects.

A further difficulty with Frazier & Rayner's approach is that they do not give an explicit algorithm for the delay strategy, so that this strategy cannot easily be tested on additional cases of lexical ambiguity.[30]

---

[30]While it is unclear from Frazier & Rayner's paper precisely what delay algorithm they have in mind (no

A final possible problem for Frazier's theory is that it makes no amends for interplay between syntax and other levels of linguistic analysis. That is, the theory of Minimal Attachment predicts that the structure with the fewest number of nodes will always be pursued by the human processor first, independent of semantic, pragmatic and contextual factors. There is currently a great deal of controversy over whether or not this prediction is correct, but I believe that the weight of the evidence is against a syntax-first approach like Minimal Attachment (see, for example, Tyler & Marslen-Wilson (1977), Marslen-Wilson & Tyler (1980, 1987), Milne (1982), Schubert (1984, 1986), Kurtzman (1985), Wilks, Huang & Fass (1985), Carlson & Tanenhaus (1988), Crain & Steedman (1985), Taraban & McClelland (1988, 1990), Altmann (1988), Altmann & Steedman (1988), Ni & Crain (1989), Stowe (1989), Trueswell, Tanenhaus & Garnsey (submitted), *cf.* Forster (1979), Frazier & Rayner (1982), Rayner, Carlson & Frazier (1983), Ferreira & Clifton (1986)). See Sections 9.3 and 9.4 for a discussion of some of this evidence.

## 2.3. The Theta Reanalysis Constraint: Pritchett (1988)

As a result of some of the difficulties with Minimal Attachment and Late Closure mentioned in Section 2.2.1, Pritchett (1988) proposes an alternative processing theory which collapses the Principles of Minimal Attachment and Late Closure into a single parsing principle:

(50) Σ-Attachment: Every principle of the Syntax attempts to be satisfied at every point during processing.

The syntactic principles that Pritchett assumes are those from Government and Binding Theory (Chomsky (1981), Chomsky (1986a)). In particular he appeals to the $\theta$-Criterion:

(51) The $\theta$-Criterion: Each argument bears one and only one $\theta$-role (thematic role) and each $\theta$-role is assigned to one and only one argument. [Chomsky (1981), p. 36]

Thus given an argument, the parser prefers to assign a thematic role to that argument so that the $\theta$-Criterion will be locally satisfied. For example, consider once again (52):

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

---

algorithm is given), one possibility is as follows:

Given a categorically ambiguous word together with the structure for the current input string, do the following:

1. Consult the phrase structure rules to see if only one of the ambiguous categories can be added to the structure assembled thus far.
2. If only one of the categories is appropriate, then build the appropriate structure. If not, then input the next word and go to step 1.

Note that this algorithm effectively consists of building all possible structures for the input string including the ambiguous word, and then seeing if there is exactly one of these structures. It seems strange to call such a strategy "a delay strategy," however, since multiple structures for the input are being constructed at some level of analysis.

There is a local ambiguity at the point of parsing the word *raced*. This verb can either attach as the matrix verb or as a modifier of the NP *the horse*. In the matrix verb attachment the NP *the horse* receives a thematic role from the verb *raced*. No such thematic role is assigned to *the horse* in the modifier attachment. Thus the matrix verb attachment is locally preferred since it better satisfies the $\theta$-Criterion than does the modifier attachment. When the word *fell* is processed, no attachments are possible and backtracking is necessary to obtain a parse for the sentence. Thus the garden-path status of (52) is predicted.

As it stands, this theory does not predict the lack of conscious garden-path effects in sentences like (53):

(53) Bill knew John liked Mary.

Consider the state of the parse of (53) when the word *John* is input. The preferred structure will have this NP attached as direct object of the verb *know* so that it can receive a thematic role. When the verb *liked* is then input, no attachments are possible to the current structure, so reanalysis is necessary to rectify the situation. Thus (53) is incorrectly predicted to induce a garden-path effect under this simple theory. In order to account for the non-garden-path status of sentences like these, Pritchett hypothesizes the existence of the Theta Reanalysis Constraint:

(54) Theta Reanalysis Constraint (TRC): Syntactic reanalysis which interprets a $\theta$-marked constituent as outside its current $\theta$-Domain is costly.

(55) $\theta$-Domain: $\alpha$ is in the $\gamma$ $\theta$-Domain of $\beta$ iff $\alpha$ receives the $\gamma$ $\theta$-role from $\beta$ or $\alpha$ is dominated by a constituent that receives the $\gamma$ $\theta$-role from $\beta$.

Consider (53) with respect to the Theta Reanalysis Constraint. At the point of parsing the word *liked*, reanalysis is necessary. The NP *John* initially receives its $\theta$-role from the verb *knew*; after reanalysis this NP receives its $\theta$-role from the verb *liked*. Pritchett hypothesizes that this reanalysis is not costly because of the Theta Reanalysis Constraint: after reanalysis the $\theta$-marked constituent *John* is still within its original $\theta$-Domain since the $\theta$-role initially assigned to *John* is now assigned to a constituent dominating *John*, the complementizer phrase *John liked*.[31]

Note that the definition of the Theta Reanalysis Constraint does not change the garden-path prediction for (52). At the parse state just before the word *fell* is input, the NP *the horse* receives the $\theta$-role AGENT from the verb *raced*. When the word *fell* is input, the NP *the horse* must be reanalyzed as THEME of *fell* rather than as AGENT of *raced*. Since this reanalysis interprets the NP *the horse* as outside its original $\theta$-Domain, this reanalysis is expensive and a garden-path effect results.

While this theory now accounts for the non-garden-path status of sentences like (53),

[31] The category complementizer phrase (CP) is the equivalent of the traditional $S'$ node (Chomsky (1986b). Furthermore, tense and agreement information are assumed to reside in the category *Infl*. The category IP (Infl phrase) is the modern equivalent of the traditional $S$ node.

Pritchett notes that it does not yet account for the lack of garden-path effect in sentences like (56):

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

In order to account for the contrast between (56) and the garden-path sentence in (52), Pritchett hypothesizes a modification to the Theta Reanalysis Constraint:

(57) Theta Reanalysis Constraint (revised): Syntactic reanalysis which interprets a $\theta$-marked constituent as outside its current $\theta$-Domain and as within an existing $\theta$-Domain of which it is not a member is costly.

The difference between (52) and (56) is derived from the fact that the verb *found* is obligatorily transitive, while the verb *raced* is optionally intransitive. According to Pritchett's analysis, the verb *found* is initially treated as the matrix verb, just as *raced* is initially analyzed as the matrix verb in (52). However, since the verb *found* is obligatorily transitive, the next word in the input string, *in*, forces reanalysis. That is, since *found* must assign accusative case, and there is no NP present to which it may assign this case, the main clause analysis of *found* is revoked in favor of the reduced relative clause reading. This reanalysis does not cause difficulty because of the way that the Theta Reanalysis Constraint has been revised.

In addition to handling many of the effects predicted by Minimal Attachment, Pritchett's theory also predicts garden-path effects in (58a) and (58b), thus partially collapsing Frazier's principles of Minimal Attachment and Late Closure:

(58)a. # While Mary was singing a song played on the radio.
b. # Without her contributions arrived infrequently.

Consider (58a). When the NP *a song* is encountered it can either attach as direct object of the verb *singing* where it would get a thematic role from *singing*, or it can attach as subject of the matrix clause, where it would receive no thematic role. The direct object attachment is thus preferred because this structure best satisfies the grammar. Furthermore, when the verb *played* is input, expensive reanalysis is necessitated, since the NP *a song* must be reanalyzed as outside of its current $\theta$-Domain, that associated with the verb *singing*. Thus garden-path effects are predicted by Pritchett's theory for both (58a) and (58b), as desired.

Furthermore, Pritchett's theory has no trouble with sentences like (59), avoiding a difficulty associated with Minimal Attachment:

(59) # Once they had the computer companies worked more efficiently.

As in (58b), the preferred attachment of the NP *companies* is as head of the NP object of the preposition *without*. Expensive reanalysis is forced by the Theta Reanalysis Constraint and a garden-path effect correctly predicted.

### 2.3.1. Problems with Pritchett's Theory

While Pritchett gives a theory of garden-path effects that is appealing 1) because of its ability to account for a large quantity of data and 2) because of its reliance on an independently motivated theory of syntax, there are a number of empirical and theoretical difficulties with his model. First of all, as is the case for the theory of Minimal Attachment and Late Closure, Pritchett assumes that syntactic analysis precedes semantic, pragmatic and contextual analysis despite the existence of much empirical evidence that contradicts this assumption. See Section 2.2.1 for a list of some of the appropriate references.[32]

Second, consider the sentences in (60), sentences whose preferred readings are predicted by Frazier's principle of Late Closure:

(60)a. Bill thought John died yesterday.
b. John figured that Sue wanted to take the cat out.

Consider (60a). No principle of the syntax is better satisfied by attaching the adverb *yesterday* to the embedded clause rather than to the matrix clause. Thus the preferred reading of (60a) is unexplained in Pritchett's theory: his theory will be forced to include a principle like Late Closure in order to account for examples like (60a). Similarly, no syntactic principle is better satisfied by associating the particle *out* with the verb *take* rather than the verb *wanted* in (60b). A principle like Late Closure seems to be needed independently.[33]

A third source of difficulty for Pritchett's theory comes from sentences like the following:

(27a) # Since she jogs a mile seems light work.

(61)a. # Since she jogs a mile seems light work.
b. # While Mary was singing a song was playing on the radio.
c. # The horse raced past the barn was falling.
d. # Without her contributions would be inadequate.
e. # Without the computer companies would be less productive.

The Theta Reanalysis Constraint as defined in (57) incorrectly predicts that each of the above sentences does not induce a garden-path effect. These sentences differ from previous examples which induce garden-path effects (*e.g.*, (52), (58) and (59)) in the type of verb which initiates reanalysis. None of the reanalysis-forcing verbs in the above sentences assigns any thematic role to its subject, so that the reanalysis does not take place in any $\theta$-Domain, thus avoiding Theta Reanalysis cost. For example, consider (61c). When the auxiliary verb

[32] However, as noted in Section 2.2.1, this issue is not clear-cut: there is still a great deal of controversy.

[33] The fact that Pritchett's theory does not handle examples like those in (60) might not be a problem, since it may be possible to incorporate a principle like Late Closure into a framework like Pritchett's. In any case, it is important to note that Pritchett's theory does *not* get all of the desired effects of Minimal Attachment and Late Closure in one principle. The examples in (60) are sentences whose preferred structures are not explained by Pritchett's theory, and thus some additional principles are needed.

*was* is encountered, reanalysis is forced. However, the auxiliary verb *was* does not have a thematic role to assign to its subject, *the horse*, so the TRC is not violated. Thus Pritchett's theory incorrectly predicts that these sentences do not cause garden-path effects.

A further difficulty for Pritchett's theory is also rooted in the revised version of the Theta Reanalysis Constraint. The revised TRC predicts that in partial sentences like those in (62) people will initially prefer the main clause reading of the ambiguous verb:

(62)a. The events recreated {were, history}...
b. The army brigade called up {was, your}...
c. The books required {appeared, special}...
d. The scenes recorded {our, were}...
e. The monkeys chased {his, were}...

If a second verb appears immediately (as in the above examples) expensive reanalysis will be necessary, since this reanalysis takes place inside another verb's $\theta$-Domain. Contrary to this prediction, Kurtzman (1985) found that subjects had no trouble with either of the listed completions for each partial sentence, indicating that both the matrix clause and reduced relative clause readings of these verbs were available at that point in parsing. This result contradicts the prediction of the revised TRC.

Another empirical difficulty arises in garden-path sentences like (63):

(63) # I convinced her children are noisy.

Consider the state of the parse as the word *children* is being processed. It can either become the head for the NP argument *her children* or it can receive the second internal thematic role from *convinced*. In order to successfully predict the garden-path in (63), it is necessary for the parser to initially treat *her children* as a single NP. However, the parsing principles given thus far do not give this prediction. If *her children* is analyzed as a single NP, then it receives a thematic role in accordance with the $\theta$-Criterion, but the verb *convinced* still has a thematic role to assign, in local violation of the $\theta$-Criterion. On the other hand, if the input string *her children* is analyzed as two NPs, then the $\theta$-Criterion is completely satisfied: all arguments receive thematic roles and all thematic roles are assigned. Thus, if only the $\theta$-Criterion is considered, exactly the wrong prediction is made.

In order to account for this discrepancy, Pritchett appeals to Case Theory. According to the Case Filter, lexical NPs need abstract Case. If *her children* is analyzed as an NP, then all NPs receive Case as desired. If, on the other hand, *her children* is two separate NPs, then only the first receives Case, and the second exists in local violation of the Case Filter. Pritchett stipulates that this local violation of the Case Filter is not permissible. As a result, the preferred analysis treats *her children* as a single NP.

Note that this solution to the problem observed in (63) is not satisfactory. There are two possible analyses for the input string *her children*. The first, which treats *her children* as a noun phrase, includes a local violation of the $\theta$-Criterion, since one thematic role is unassigned. The second, which treats *her children* as two separate arguments of *convinced*, contains a local violation of the Case Filter, since the NP *children* does not receive Case.[34] Pritchett simply stipulates, without independent evidence, that one local violation of the Case Filter is worse than one local violation of the $\theta$-Criterion. This is unsatisfactory.

If Case Theory must be locally satisfied, as Pritchett suggests, a further difficulty arises in sentences like (53):

(53) Bill knew John liked Mary.

If the Theta Reanalysis Constraint is not to be stipulative, then it should follow from (50). Since Pritchett appeals to Case Theory in many of his garden-path derivations, there should be a corresponding Case Reanalysis Constraint. However there cannot be such a constraint because of examples like (53). If the Case Filter is to be locally satisfied, then the preferred attachment of the NP *John* is as direct object of the verb *knew* so that it receives accusative Case. Case reanalysis is necessary when the verb *liked* is encountered: the noun *John* now receives nominative Case. Since this sentence is not difficult to process, this reanalysis should not be costly. Thus either (53) violates the Case Reanalysis Constraint or there is no Case Reanalysis Constraint and hence the Theta Reanalysis Constraint is stipulative.

Furthermore, Pritchett's theory has difficulty accounting for the ease in which people process many lexical ambiguities. Some of the relevant examples are given below:

(64)a. The warehouse fires numerous employees each year.
b. The warehouse fires harm some employees each year.

(65)a. I gave her earrings on her birthday.
b. I gave her earrings to Sally.

(66)a. Have the boys take the test today.
b. Have the boys taken the test today?

Consider (64a) and (64b). The first three words of these sentences are ambiguous between noun phrase and sentential readings. Pritchett's theory predicts that the sentential reading will be preferred since thematic roles are assigned to all NPs in this representation, while an NP is left without a thematic role in the other possible structure. Reanalysis which violates the TRC ensues and a garden-path effect is incorrectly predicted for (64b). Similarly, since Pritchett assumes a serial model, at most one representation can be maintained for the local ambiguities in (66a), (66b), (65a) and (65b). Reanalysis results in order to get the other reading, and since this reanalysis violates the TRC, garden-path effects are incorrectly predicted for at least half of these sentences.

An additional problem with Pritchett's theory is given by the lack of difficulty that people have with examples like (67):

[34] Crucially note that this NP can receive Case from a following tensed verb. In fact, this is exactly the situation that occurs in (63).

(67) The patient persuaded the doctor that was having trouble to leave.

Pritchett's theory predicts that people will initially analyze the complementizer *that* as an argument to the verb *persuaded*. When this analysis turns out to be incorrect, reanalysis needs to take place. This reanalysis is predicted to be conscious since it violates the TRC. However, no conscious garden-path effect occurs in (67), contrary to Pritchett's prediction.[35]

Sentences (68a) and (68b), which pose a problem for theory of Minimal Attachment, also gives Pritchett's theory difficulty:

(68)a. The report that the president sent to us helped us make the decision.
b. The report that the president sent the troops into combat depressed me.

As noted earlier, the sentences in (68) are both ambiguous at the point of processing the complementizer *that*. Pritchett's theory predicts that the argument attachment should be locally preferred, so that backtracking is necessary in the parse of (68a). Furthermore, this backtracking violates the TRC and is thus predicted to be conscious. However, no conscious garden-path effect occurs in the processing of (68a), contrary to prediction.

Pritchett's model also under-predicts garden-path effects, as is demonstrated by the examples in (69):

(69)a. # I believe that John smokes annoys Mary.
b. # The populace knows that Dan is stupid is obvious.
c. # The students thought that Rick ate a sandwich larger than his head was impressive.

Consider (69a). The complementizer phrase *that John smokes* is initially attached as complement of the verb *believe*. When the verb *annoys* is encountered, reanalysis takes place. But this reanalysis does not violate the Theta Reanalysis Constraint in the same way that the reanalysis in (53) does not violate the TRC. Thus Pritchett's theory predicts that the sentences in (69) are as easy to process as (53). This prediction is obviously wrong: (53) gives readers no difficulty, while the sentences in (69) cause severe garden-path effects. Note that Minimal Attachment makes the correct prediction for the sentences in (69), while it makes the wrong prediction for (53), exactly the opposite behavior of Pritchett's theory.

## 2.4. Deterministic Parsing with Lookahead: Marcus (1980)

One alternative to the serial hypothesis is provided by Marcus (1980). In order to explain the rapidity with which people process natural language, Marcus proposed that the human parser operates *deterministically*. A processor is deterministic if 1) any structure that it builds is part of the final representation for the input and 2) no temporary structures are encoded within internal machine states. While forcing a parser to build structure deter-

[35] Note that it is not possible to explain the lack of difficulty in (67) by appeal to the short distance over which reanalysis takes place. See footnote 25 for a similar argument with respect to the theory of Minimal Attachment.

ministically without looking ahead in the input is extremely efficient (Knuth (1965)) and therefore desirable, Marcus noted that many sentences are locally ambiguous but yet give the human processor no difficulty:

(66a) Have the boys take the test today.
(66b) Have the boys taken the test today?

(70)a. Is the block sitting in the box?
 b. Is the block sitting in the box red?

Consider the sentences in (70a). In (70a) the verb *is* functions as an auxiliary verb to the verb *sitting*. On the other hand, in (70b) the verb *is* functions as the main verb of the sentence, linking the subject *the block sitting in the box* to the adjectival predicate *red*. Neither of the sentences in (70) is difficult to process, thus ruling out the possibility that a deterministic parser can successfully model human linguistic processing without lookahead. As a result, Marcus hypothesized that the human parser can look ahead up to three constituents in the input string. A three constituent lookahead buffer allows the parser to handle sentences like those in (66) and (70), but is finite in size so that parsing efficiency is still linear in the input length (Knuth (1965)). Furthermore, Marcus hypothesized that his model would explain the existence of garden-path effects.[36] That is, if a sentence cannot be processed deterministically given a lookahead buffer of three constituents, it will be predicted to cause conscious difficulty: a garden-path effect.

Marcus' model consisted of a stack along with a three-cell buffer. Structures were built by consulting rules in an external grammar. The rules could refer to structures in one of four locations: the three buffer cells and the top of the stack.

Consider Marcus' explanation of (71), given his partial grammar for English:

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

After the NP *the barn* has been processed, the stack contains an NP representation for *the horse* attached to a predicted sentential node. The three cell lookahead buffer contains the following three constituents :

$[_V$ raced $]$ $[_{Prep}$ past $]$ $[_{NP}$ the barn $]$

Crucially, the verb *fell* is not yet visible because of the limited size of the buffer. If it were visible, the grammar would attach it as the main verb of the sentence, thus forcing the words *raced past the barn* to attach as a modifier of the NP *the horse*. However, since the verb *fell* is not visible, the grammar attaches *raced* as the main verb of the sentence, and a garden-path effect results when the verb *fell* can't attach anywhere in the current structure.

[36]Marcus also tried to derive Subjacency effects from the architecture of his parser (Marcus (1980), Berwick & Weinberg (1984, 1985)). See, however, Fodor (1985) and Smith (1986) for criticisms of this attempt.

### 2.4.1. Problems with the Marcus Parser

While Marcus' abstract proposal that the human parser can look ahead a finite distance is plausible, the particular formulation of the lookahead that he proposes has many empirical difficulties. First consider reduced relative clause/main clause ambiguities like that in (71). A major difficulty with Marcus' treatment of sentences like (71) is that it crucially relies on the fact that *past the barn* is not analyzed in the buffer as a prepositional phrase (Church (1980)). That is, if the rules of the grammar allowed the PP *past the barn* to be built before the verb *fell* were input, then the parser could disambiguate the sentence. Marcus explicitly argues for a noun phrase preprocessor on the grounds that the initial components of NPs – usually determiners in Marcus' view – are clear-cut. Such an argument, whatever its merits, clearly applies to prepositional phrases if it applies to noun phrases, since PPs are initiated by prepositions: a closed class group. Thus the rules in Marcus' grammar should allow PPs to be preprocessed, thus disambiguating sentences like (71) within the three buffer cell limit.

Even if we accept the stipulation that NPs are preprocessed but PPs are not, Marcus' account of the garden-path effect in (71) is unsatisfactory. Marcus' prediction of a garden-path effect in (71) relies on the fact that there are three constituents between the subject NP *the horse* and the matrix verb *fell*. Thus having less than three constituents should not result in a garden-path effect. Hence (72) is incorrectly predicted to be unproblematic for the human parser (Pritchett (1988)):

(72) # The horse raced quickly fell.

That is, in the processing of (72) there will be a parse state in which the verb *raced* (as yet undetermined to be passive participle or past tense), the adverb *quickly* and the verb *fell* are all present in the buffer, so that the grammar can determine that the passive participle reading of *raced* is the appropriate one. Thus Marcus' model incorrectly predicts that (72) does not induce a garden-path effect.

There exist numerous other syntactic garden-path effects that the three cell buffer of the Marcus parser does not predict, among them those in (73) (Briscoe (1983)):[37]

(73)a. # The patient persuaded the doctor that he was having trouble with to leave.
b. # Without her contributions would be inadequate.
c. # While Mary was singing a song played on the radio.
d. # I gave the boy the dog bit a bandage

Consider for example (73c). Since the buffer is three cells in length, it can easily hold representations for the NP *a song* as well as the verb *played* while the top of the stack contains an $S'$ representation of *while Mary was singing*. As a result, the Marcus parser has no non-stipulative way of predicting the garden-path effect in (73c). Similarly, the other sentences in (73) can also be disambiguated within the three buffer cell limit.

[37]While Marcus did not claim that all garden-path effects would be predicted by the three cell lookahead, additional mechanisms (rule ordering *etc.*) would be necessary in order to predict these effects.

The Marcus parser not only does not predict garden-path effects when it should, it also predicts some garden-path effects that it do not occur. Consider (74) (Pritchett (1988)):

(74) I know my aunt from Peoria died.

In order to account for the garden-path effect in (71), Marcus assumes that prepositional phrases are not preprocessed with noun phrases. As a result, the three cells of the buffer are occupied at the parse state immediately before the verb *died* is to be input. In order to make room for *died*, the NP *my aunt* is attached as direct object of the verb *know* on the stack. A garden-path effect is thus predicted since backtracking will be necessary to rectify the parse.

Furthermore, the Marcus parser predicts that sentences with easily recognizable global ambiguities cannot exist, contrary to fact (Fodor (1985)). Consider, for example, (75):

(75) I saw her duck.

Sentence (75) has two readings, one in which *duck* is a noun and one in which *duck* is a verb. People easily recognize both of these readings. However, the Marcus parser predicts that one of the two readings is hard to obtain, an undesirable consequence of the theory.

### 2.4.2. Description Theory

In order to avoid some of the empirical difficulties of the original Marcus parser and yet maintain the determinism hypothesis, Marcus, Hindle & Fleck (1983) propose an updated model which used the same three buffer cell data structures but explains difficulty in terms of *Description Theory* or D-theory.[38] Following Lasnik & Kupin (1977), Marcus, Hindle & Fleck propose that the output of a parse should not consist of a tree structure *per se*, as is commonly assumed; rather the output should consist of statements about the appropriate dominance relationships among categories. An example will make their approach clearer. Consider the sentence in (76) along with the tree structure that is normally associated with it in (77):

(76) John sees Mary.

(77) $[_{S_1}$ $[_{NP_1}$ John ] $[_{VP_1}$ $[_{V_1}$ $[_{NP_1}$ Mary ]]]]

Without actually building the structure in (77), it is possible to list all the dominance relationships among the categories in (77), assuming the linear precedence of words in (76). Some of the dominance relationships are given in (78):

[38]See also Weinberg (1988) for a model using D-theory without the Marcus parser as its underlying structure builder.

(78)a. $NP_1$ dominates *John*

b. $S_1$ dominates $NP_1$

c. $S_1$ dominates $VP_1$

d. $VP_1$ dominates $V_1$

e. $V_1$ dominates *sees*

f. $VP_1$ dominates $NP_2$

g. $NP_2$ dominates *Mary*

Given the parse output in terms of dominance relationships, Marcus, Hindle & Fleck propose that parsing difficulty occurs when, during the course of a parse, one or more dominance relations has to be removed from the current set. This assumption allows them to explain the lack of garden-path effect in (74):

(74) I know my aunt from Peoria died.

Despite the fact that the parser initially misattaches the NP *my aunt* as direct object of the verb *know*, reanalysis does not require any dominance relationships to be removed from the parse output. That is, the parser initially posits the following relationships, among others:

(79)a. $NP_2$ dominates *my aunt*

b. $VP_1$ dominates $NP_2$

When the word *died* is input, it becomes necessary to add the following further dominance relationships:

(80)a. $VP_1$ dominates $S_2$

b. $S_2$ dominates $NP_2$

Since the tree structure is never actually built, no structure is destroyed. Dominance relations are only added, so no difficulty is predicted with (74), as desired.

However, in spite of this improvement over the theory of Marcus (1980), the theory of processing breakdown provided by Marcus, Hindle & Fleck (1983) is little better at predicting garden-path effects than is Marcus' original parser. Since the updated model still uses the three cell look-ahead, it incorrectly predicts that none of sentences (72) and (73) induce garden-path effects, in the same way as did the Marcus parser. Thus until these difficulties are addressed, the model of Marcus, Hindle & Fleck must also be rejected as a model of parsing difficulty.

## 2.5. Lexical Preference: Ford, Bresnan & Kaplan (1982)

A lexically-based theory of attachment preferences and garden-path effects is given by Ford, Bresnan & Kaplan (1982). In this paper Ford, Bresnan & Kaplan (FB&K) note that two words with the same subcategorization patterns may differ in argument attachment preferences. For example, the verbs *object* and *signal* each take optional prepositional phrase and sentential complements, but these verbs differ in their attachment preferences, as is demonstrated by the sentences in (81):

(81)a. The tourists objected to the guide that they couldn't hear.

b. The tourists signaled to the guide that they couldn't hear.

In the preferred readings of each of (81a) and (81b) the PP *to the guide* attaches as an argument of the matrix verb: *objected* in (81a) and *signaled* in (81b). However, the $S'$ structure *that they couldn't hear* prefers to be attached to the matrix verb *signaled* in (81b), while it prefers to be attached as modifier of the NP *the guide* in (81a) where the matrix verb is *objected*. Since these two verbs have the same subcategorization patterns, structural preference heuristics such as Minimal Attachment and Late Closure by themselves cannot predict the patterns observed in (81). That is, a purely structural analysis of the preferences in (81) makes the same predictions for each sentence, and thus is insufficient.

As a result, FB&K propose the principles of Lexical Preference and Final Arguments among their parsing principles in a serial processing framework. These principles are based upon the assumption that the lexical entries for a word are ranked, where the most preferred of these entries is the known as the *strongest lexical form* for a word.

(82) Lexical Preference: If a set of alternatives has been reached in the expansion of a phrase structure rule, give priority to the alternatives that are coherent with the strongest lexical form of the predicate.

(83) Final Arguments: Give low priority to attaching to a phrase the final argument of the strongest lexical form of that phrase and to attaching any elements subsequent to the final element. Low priority is defined here with respect to ther options that arise at the end position of the element whose attachment is to be delayed.

These two principles interact within the Lexical Functional Grammar framework (Bresnan (1982)) to give the result that preference is given to structures that result from attachment to stronger lexical forms. In order to account for the preference in (81b), it is assumed that the strongest lexical form of the verb *signal* in a neutral context is one that subcategorizes for a subject, a prepositional phrase object and a sentential complement. Thus when processing (81b), there is a preference to attach the $S'$ *that they couldn't hear* to the verb *signaled* in order to maintain the strongest lexical form of *signal*. Furthermore, it is assumed that the strongest lexical form of the verb *object* in a neutral context is one that subcategorizes for a subject and a prepositional phrase object, but no sentential complement. Thus, because of the Principle of Final Arguments, the prepositional phrase *to the guide* delays its attachment to the verb *objected* as long as possible. As a result, the $S'$ *that they couldn't hear* attaches to the NP *the guide*, and the preferred reading in (81a) is explained.[39]

[39]Note that the lexical entry for *object* that subcategorizes for a PP complement but no sentential complement has a different meaning from the lexical entry that subcategorizes for both a prepositional phrase and a sentential complement. To see this, consider the sentences in (84) and their respective paraphrases:

These two principles explain the difference between the sentences in (85) in a similar manner:

(85)a. The woman positioned the dress on that rack.

b. The woman wanted the dress on that rack.

The preferred reading of (85a) links the PP *on that rack* to the verb *positioned*, while the preferred reading of (85b) links the PP *on that rack* to the NP *the dress*. FB&K explain the existence of these varying preferences by appeal to lexical preference: The strongest lexical form of the verb *position* is one that subcategorizes for a subject, a direct object and an oblique PP, while the strongest lexical form of *want* is one that subcategorizes for only a subject and a direct object. Thus the preferred reading in (85a) is explained since the PP *on that rack* prefers to attach to the strongest lexical form of the verb. And the Principle of Final Arguments explains the preference in (85b): the NP *the dress* delays its attachment to the verb *wanted* as long as possible, thus blocking the attachment of the PP *on that rack* to any node higher than the NP *the dress*.

FB&K suggest that their approach to attachment preferences can be extended in order to explain the existence of strong garden-path effects with the addition of one assumption: a strong garden-path effect occurs when a previously analyzed word needs to be morphosyntactically reanalyzed. That is, given their lexical and syntactic parsing principles, if it happens that reanalysis is necessary in order to successfully parse a sentence, then that reanalysis is unproblematic except when the reanalysis includes morphosyntactic reanalysis of a word. This assumption explains the difficulty in each of the following garden-path sentences:

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

(87) # The boy got fat melted.

In (86), the simple past tense reading of the verb *raced* is initially preferred by FB&K's parsing principles,[40] but this structure is not compatible with the rest of the sentence to follow. Reanalysis is needed when the verb *fell* is encountered. In order to get an appropriate reading for (86), morphosyntactic reanalysis of the verb *raced* is needed: this word must be reanalyzed as a passive participle. Thus FB&K correctly predict a strong garden-path effect in this case.

The garden-path effect in (87) is explained in a similar manner. The word *fat* is initially categorized as an adjective, but this categorization turns out to be incorrect and must be

(84) a. The tourists objected to the guide.

"The tourists had objections to the guide."

b. # The tourists objected to the guide that they were tired.

"The tourists complained to the guide that they were tired."

That is, the word *objected* is paraphrased as *had objections* in (84a), while the same word is paraphrased as *complained* in (84b), two related but distinct meanings for *objected*. Note that the lexical entry for *object* in (84a) ("have objections") does not optionally take a sentential complement.

[40]This preference comes about as a result of the principle of Invoked Attachment, a principle related to Minimal Attachment. See Ford, Bresnan & Kaplan (1982) for a description of this principle.

reanalyzed. Since the reanalysis involves changing the category of the word *fat*, a strong garden-path effect is predicted, as desired.

FB&K note that this analysis predicts that a sentence like (88) will not cause the same level of garden-path effect as in (86) and (87):

(88) # The tourists signaled to the guide that they didn't like.

In (88) the $S'$ structure *that they didn't like* will be initially analyzed as a complement of the verb *signaled* in order to satisfy that verb's strongest lexical form. When this attachment is not compatible with the rest of the sentence, it is reanalyzed as a modifier of the NP *the guide*. FB&K state that since the garden-path effect that people experience in (88) is not nearly as strong as those in (86) and (87), their model seems to make the right predictions.

### 2.5.1. Problems with the Ford, Bresnan & Kaplan (1982) Model

While FBK's paper was very important in that it pointed out the need to consider lexical differences when computing parsing preferences, the theory of preferred readings and garden-path effects that is put forward is not adequate because of a number of empirical shortcomings.

First, given that the strongest lexical form of the verb *want* subcategorizes for a subject and an object only, the Principle of Final Arguments makes the wrong prediction in examples like (89) (Schubert (1984), Wilks, Huang & Fass (1985)):

(89) The woman wanted the dress for Mary.

The preferred reading of (89) has the PP *for Mary* attached to the main verb *wanted*. However, the Principle of Final Arguments predicts that this PP should prefer to attach to the NP *the dress*. That is, since the final argument of *want* will be satisfied if the NP *the dress* is attached, this attachment is delayed as long as possible, thus forcing the PP *for Mary* to attach low.

Furthermore, there are many difficulties with FB&K's theory of garden-path effects. First of all, their explanation of the garden-path effect in (86) gives no way to account for the lack of garden-path effects in sentences like (90):

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

FB&K's theory predicts that the word *found* will first be analyzed as a simple past tense verb, the matrix verb of the sentence, in the same way that *raced* is initially analyzed as the matrix verb of (86). When this analysis turns out to be incorrect, backtracking is necessary and the word *found* is reanalyzed as a passive participle. Since this reanalysis is morphosyntactic, FB&K's theory predicts a strong garden-path effect, just as in (86). However, there is in fact no conscious garden-path effect in (90), contrary to FB&K's predictions.

Similarly, given a lexical ambiguity that can be resolved in more than one way, the model of garden-path phenomena proposed by FB&K predicts that one of the resolutions will cause severe processing difficulty. This prediction is clearly incorrect, as the following examples illustrate:

(91)a. The warehouse fires numerous employees each year.
b. The warehouse fires harm some employees each year.

(92)a. John likes fat on his steak.
b. John likes fat cats.
c. John saw the light.
d. John saw the light beer.

Consider (92a). If the word *fat* is initially categorized as an adjective as FB&K suggest in their explanation of the garden-path effect in (87), then sentence (92a) is incorrectly predicted to induce a garden-path effect. If, on the other hand, *fat* is initially categorized as a noun, then (92b) is incorrectly predicted to induce a garden-path effect. Thus FB&K's model makes undesirable predictions with respect to many lexical ambiguities.

FB&K's model not only predicts many non-garden-path sentences to be garden-path sentences, it also does not predict the garden-path status of many garden-path sentences. Recall that FB&K assume that (88) does not induce a strong garden-path effect:

(88) # The tourists signaled to the guide that they didn't like.

While it may be that the garden-path effect in (88) is not as strong as it is in (86), (88) still requires conscious backtracking, a fact that remains unexplained in the FB&K model. Furthermore, FB&K predict no difference between sentences like (88) and those to follow:

(93)a. Bill knew John.
b. Bill knew John liked Mary.
c. The city council argued the mayor's position.
d. The city council argued the mayor's position was incorrect.

FB&K's theory predicts no difficulty with any of the sentences in (93) since any reanalysis that takes place is not morphosyntactic reanalysis. This prediction is correct, but the contrast between the sentences in (93) and (88) remains unexplained.

Furthermore, no morphosyntactic reanalysis takes place in any of the following examples, yet each induces a conscious garden-path effect:

(94)a. # The patient persuaded the doctor that he was having trouble with to leave.
b. # While Mary was singing a song played on the radio.
c. # I believe that John smokes annoys Mary.

For example, the garden-path effect in (94c) is certainly as strong as that in (87). This fact is left unexplained by the FB&K model.

As a result of the many empirical shortcomings noted here, the FB&K model cannot be adopted as an explanation of garden-path effects and parsing preferences. Nevertheless, the FB&K model is important since it was one of the first to point out some of the effects on processing of multiple lexical entries for the same word. See Section 9.2 for a description of how lexical preferences are handled in the system proposed here.[41]

### 2.6. Semantics and Context Based Approaches: Crain, Steedman & Altmann

A large number of researchers have observed that semantics, pragmatics and context have a great influence on the human parser's choices at points of ambiguity (see, for example, Tyler & Marslen-Wilson (1977), Marslen-Wilson & Tyler (1980, 1987), Milne (1982), Schubert (1984, 1986), Kurtzman (1985), Wilks, Huang & Fass (1985), Carlson & Tanenhaus (1988), Crain & Steedman (1985), Taraban & McClelland (1988, 1990), Altmann (1988), Altmann & Steedman (1988), Ni & Crain (1989), Stowe (1989), Trueswell, Tanenhaus & Garnsey (submitted). In this section I will survey aspects of the approach due to Crain, Steedman and Altmann (Crain & Steedman (1985), Altmann & Steedman (1988), Altmann (1988)). Crain & Steedman (1985) give a partial theory of parsing preferences based upon semantics, pragmatics and context. They propose that the Principle of A Priori Plausibility guides preferences:[42]

(95) The Principle of A Priori Plausibility:
If a reading is more plausible in terms either of general knowledge about the world, or of specific knowledge about the universe of discourse, then, other things being equal, it will be favored over one that is not. [Crain & Steedman (1985), p. 330]

According to Crain & Steedman, the Principle of A Priori Plausibility explains the difference between the sentences in (96):[43]

(96)a. The teachers taught by the Berlitz method passed the test.
b. The children taught by the Berlitz method passed the test.

Since the NP *the teachers* is plausible as the subject of a verb like *teach* and not so plausible as object of *teach*, Crain & Steedman hypothesize that the preferred attachment of *taught* will be as matrix clause in (96a). Similarly, since the NP *the children* is not plausible as the subject of the verb *teach* while it is plausible as the object of this verb, Crain & Steedman hypothesize that the preferred attachment of *taught* in (96b) will be as a relative clause modifier. In experiments performed by Crain (1980), a significant difference in

[41]For further information regarding lexical effects during parsing, see Clifton, Frazier & Connine (1984), Mitchell & Holmes (1985), and Mitchell (1989) as well as the references cited in these papers.

[42]Others in the list of references given above have similar approaches and data. See Section 9 for a description of some of the relevant data and results.

[43]Bever (1970) was perhaps the first to note differences like those found in (96).

grammaticality judgment was found between the two kinds of sentences, thus giving support to their Principle of A Priori Plausibility.[44]

Furthermore, Crain & Steedman (1985) offer the Principle of Parsimony as a special case of the Principle of A Priori Plausibility with respect to discourse:

(97) The Principle of Parsimony:
If there is a reading that carries fewer unsatisfied but consistent presuppositions or entailments than any other, then, other criteria of plausibility being equal, that reading will be adopted as most plausible by the hearer, and the presuppositions in question will be incorporated in his or her model. [Crain & Steedman (1985), p. 333]

The Principle of Parsimony gives a metric for determining which readings are most plausible in terms of the current discourse: the ones that violate the fewest presuppositions or entailments. Crain & Steedman report experiments that confirm their intuition that fewer violated presuppositions lead to easier processing.[45] For example, in one experiment, sentences like (99a) and (99b) were placed in two different contexts: one following (98a) and the other following (98b):[46]

(98)a. A psychologist was counselling a man and a woman. He was worried about one of them but not about the other.
b. A psychologist was counselling two woman. He was worried about one of them but not about the other.

(99)a. The psychologist told the woman that he was having trouble with her husband.
b. The psychologist told the woman that he was having trouble with to leave her husband.

Both sentences in (99) are ambiguous up to the point of parsing the preposition *with*: the subordinate clause *that he was having trouble with* may attach as a complement of the verb *told*, or it may attach as a restrictive relative clause modifier of the NP *the woman*. (99a) is syntactically disambiguated in favor of the complement attachment while (99b) is syntactically disambiguated in favor of the relative clause modifier attachment.

With no preceding context, (99a) is easy to process while (99b) induces a garden-path effect. Crain & Steedman explain this effect in terms of the Principle of Parsimony. When a definite noun phrase is mentioned in the null context, its referrent's existence is presupposed

[44]See Trueswell *et al* (submitted) for on-line support of the hypothesis that plausibility affects attachment preferences. However, also see Rayner, Carlson & Frazier (1983) and Ferreira & Clifton (1986) for evidence contradicting Crain & Steedman's results. See Section 9 for a description of some of the relevant data and results.

[45]See Kurtzman (1985) for results of a similar study. See also Altmann & Steedman (1988), Altmann (1988) and Ni & Crain (1989) for closely related arguments and data. See Ferreira & Clifton (1986) and Clifton & Ferreira (1989) for opposing arguments and evidence.

[46]This experiment was originally presented in Crain (1980), but slight problems with the materials were noted by Altmann (1988). The data given here are from Altmann's study.

in the mental model of those involved with the current discourse (*cf.* Johnson-Laird (1983)). Furthermore, a definite noun phrase with a restrictive relative clause modifier presupposes the existence of other entities also in the current discourse having the property associated with the head noun of the NP in question, none of which have the property associated with the relative clause modifier. In particular, the mental model resulting from the restrictive relative clause attachment of the sentential phrase *that he was having trouble with* to the NP *the woman* includes the following presuppositions: 1) that some set of women is already in focus in the current model; 2) that one member of this set has the property that the psychologist was having trouble with this woman; and 3) that no other members of the set of women have this property. Note that none of these presuppositions are present in the mental model resulting from the argument attachment of the sentential phrase to the verb *told*. Thus the Principle of Parsimony explains the preference for the argument attachment since fewer presuppositions are involved in this attachment than in the restrictive relative clause attachment.

Furthermore, following a context like (98b), in which there are two women in the discourse, there is no garden-path effect in the processing of (99b) while there is a garden-path effect in the processing of (99a) as predicted by the Principle of Parsimony (*cf.* Kurtzman (1985)). In particular, Crain & Steedman found that sentences like (99b) were judged grammatical more often following their supporting contexts (sentence (98b)) than following non-supporting contexts (sentence (98a)). Similarly sentences like (99a) were judged grammatical more often following their supporting contexts (sentence (98a)) than following their non-supporting contexts (sentence (98b)).

### 2.6.1. Insufficiencies of the Crain, Steedman & Altmann Model

While Crain & Steedman's Principle of A Priori Plausibility successfully predicts a number of classes of attachment preferences, there are many attachment preferences that their theory does not handle as it stands. For example, consider the difference between (100) and (101):

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

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

While (100) induces a strong garden-path effect, (101) does not. No account of this difference can be made based on context since there is no context preceding either sentence. Furthermore, no account of the difference can be made based on plausibility, since all possible attachments at the point of local ambiguity are perfectly plausible. The crucial distinction between the two sentences is that the verb *raced* is optionally transitive or intransitive while the verb *found* is obligatorily transitive: a syntactic distinction (Pritchett (1988)). No theory of parser preferences based solely on pragmatics and context will be able to account for this distinction.[47]

[47] Some researchers refer to this kind of distinction as semantic rather than syntactic. The terminology is irrelevant to the fact that this distinction is not due to pragmatics or context.

The garden-path effects in (102) are also unexplained by theories of pragmatics and context:

(102) a. # That coffee tastes terrible surprised John.
b. # I believe that John smokes annoys Mary.
c. # Without her contributions would be inadequate.
d. # I convinced her children are noisy.

For example, consider (102a). As noted in Kurtzman (1985), there is no reason to assume that the demonstrative reading of *that coffee* is associated with fewer presuppositions than the incomplete sentential reading of the same string: if anything the opposite should be true since the demonstrative reading of *that coffee* presupposes some coffee in the current discourse. Neither is there any reason to assume that the demonstrative reading is somehow more plausible than the sentential reading. Nonetheless, there is a strong preference for the demonstrative reading of the input string *that coffee* which results in a garden-path effect for this sentence. Thus Crain & Steedman's principles have no explanation for this effect. Furthermore, none of the other garden-path effects in (102) are explained by Crain & Steedman's parser preference principles.

Moreover, while Crain & Steedman's Principle of Parsimony is a possible explanation for a number of argument attachment preferences, especially ones in which the alternative possible attachment involves a restrictive relative clause, it makes incorrect predictions with respect to the following data:

(103) a. Is the block sitting in the box?
b. Is the block sitting in the box red?

(104) The patient persuaded the doctor that was having trouble to leave.

(105) a. The report that the president sent to us helped us make the decision.
b. The report that the president sent the troops into combat depressed me.

Consider the sentences in (103). Crain & Steedman's Principle of Parsimony predicts that the verb *sitting* should be attached as the matrix verb of the sentence rather than as a modifier of *the block* since fewer presuppositions are involved in this attachment. Thus Crain & Steedman's model predicts difficulty with (103b), in which the modifier attachment is pursued. However, unlike examples that Crain & Steedman give, there is no garden-path effect in the processing of (103b), contrary to predictions.[48] Similarly, the lack of garden-path effects in (104) and (105a) come as surprises to the Crain & Steedman model.

---

[48]Note that while the modifier attachment of *sitting in the box* to *the block* in (103a) and (103b) is not necessarily restrictive modification, there are still fewer presuppositions associated with the argument attachment. Presumably Crain & Steedman would argue that the difference in number of presuppositions would explain the garden-path effect in (106), which also involves an ambiguity in which the modifier attachment need not be restrictive:

(106) # I put the candy on the table into my mouth.

Hence Crain & Steedman's theory of attachment preferences is incomplete with respect to what are normally considered syntactic effects. Thus while their research correctly points out that pragmatics and context must be considered when constructing a theory of attachment preferences, it cannot be concluded that syntactic and lexical effects are nonexistent.

## 3. Previous Work: Processing Overload

In this chapter, I outline some previous approaches to determining an appropriate syntactic complexity metric that predicts gradations in acceptability. First Yngve's (1960) top-down depth-first complexity metric is discussed in Section 3.1. The second section gives Chomsky & Miller's (1963) objections to Yngve's approach along with two other metrics that they suggest for calculating sentential complexity. Fodor & Garrett's (1967) metric consisting of the ratio of words to sentential nodes is evaluated in Section 3.5. Left-corner parsing with a phrase structure rule based complexity metric is discussed in Section 3.3. Kimball's (1973) Principle of Two Sentences is given in Section 3.6. Koster's grammatical account of multiply embedded sentential subjects is discussed and evaluated in Section 3.7. Section 3.8 gives the explanation of processing overload due to Frazier & Fodor (1978). Finally, Section 3.9 outlines and evaluates Frazier's (1985) syntactic complexity metric.[49]

### 3.1. Yngve's Complexity Metric

Yngve (1960) gives a theory of sentence production that attempts to account for the complexity of multiply center-embedded structures in terms of a metric based upon a set of phrase structure rules and a stack. This model assumes that sentences are produced in a top-down depth-first manner, starting at a sentential node and recursively expanding the rules below it from left to right. Yngve proposed that the measure of complexity for the production of a node in a tree structure is given by the total number of categories that the generator has to keep track of at the point in generating the required node. This complexity metric reduces to the number of categories which 1) are on the right hand sides of incompletely expanded rules and 2) have not yet been found.[50] For example, consider the following simple context-free phrase structure grammar with respect to the sentence *the man saw the boy*:

S → NP VP
VP → V NP
NP → Det N
Det → the
N → boy, man
V → saw

[49] One metric that is not covered in this chapter is the Derivational Theory of Complexity (DTC) put forward by George Miller (see *e.g.*, Miller & Isard (1963)). This metric measures complexity in terms of the number of transformations needed to form the input string from its deep structure. There currently exist many critiques of the DTC in the psycholinguistic literature. See, for example, Garnham (1985) and the references cited there.

[50] In fact, the complexity metric should also count the category of the structure being produced, since this category must also be held in memory. However, since there is always exactly one of these categories in a top-down algorithm, the order of complexity of a given node is unchanged.

In order to generate the required input string, the category S is produced first. Then the category NP is generated since it is the leftmost category of the rule expanding the S node. At this point the category VP goes into the temporary memory store (a stack) to be expanded later. The category Det is generated next since it is the leftmost category of the NP expansion rule. The category N now goes into the temporary memory store, making a total of two categories in the temporary store. The production continues in this fashion until the whole target string is generated. The complexity, in memory units, associated with each of the nodes in the tree structure for the sentence *the man saw the boy* is given in Figure 1. As can be seen in this figure, the complexity associated with the first instance of the category Det is two memory units, These two units correspond to two categories, the category VP from the S rule and the category N from the NP rule, which are required by the grammar but are not yet present.

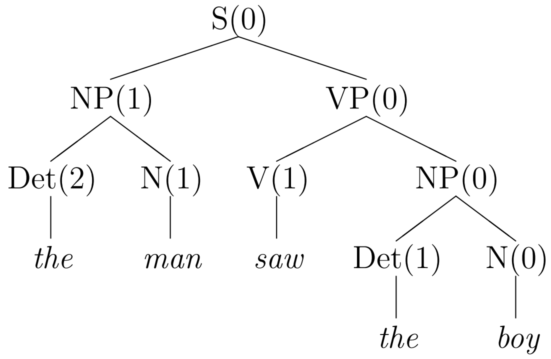


Figure 1: Yngve's top-down complexity measure for each of the nonterminals in the sentence *the man saw the boy.*

Note that while Yngve considers only generation, his complexity metric is also appropriate for a top-down processing algorithm. I will thus be considering his complexity metric for processing as well as production.[51]

Following Miller (1956), Yngve proposes that the maximal complexity that may occur in any natural language sentence is $7 \pm 2$ memory units. This limit appears to work well for right-branching structures. For example, the following right-branching sentence is not difficult to understand, despite its length (from Kimball (1973)):

(107) The dog saw the cat that chased the mouse into the house that Jack built.

Yngve's top-down complexity measures with respect to a context-free grammar are given for (107) in Figure 2. The complexity of the right-branching structure in Figure 2 is never larger than two of Yngve's memory units, and thus he correctly predicts that the structure of this sentence gives people little difficulty.

[51]See Kuno & Oettinger (1962) in which the top-down syntactic processing is first proposed for natural language.

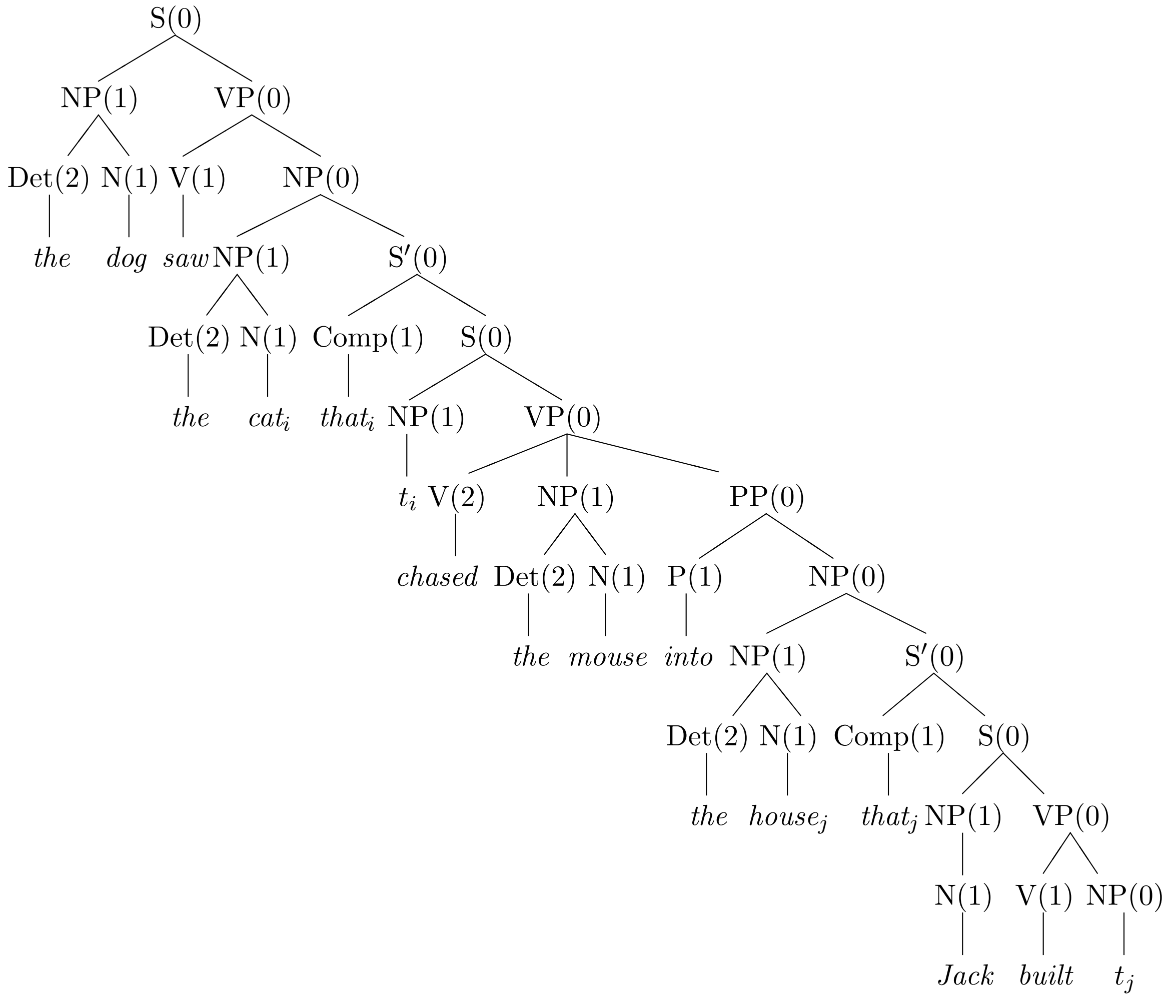


Figure 2: Yngve's top-down complexity measure for each of the nonterminals in the sentence *the dog saw the cat that chased the mouse into the house that Jack built.*

Furthermore, Yngve's complexity measure predicts a distinction between right-branching sentences and center-embedded ones as desired. For example, consider (108) along with its structure and complexity measurements in Figure 3.

(108) # The man that the woman that the dog bit saw likes fish.

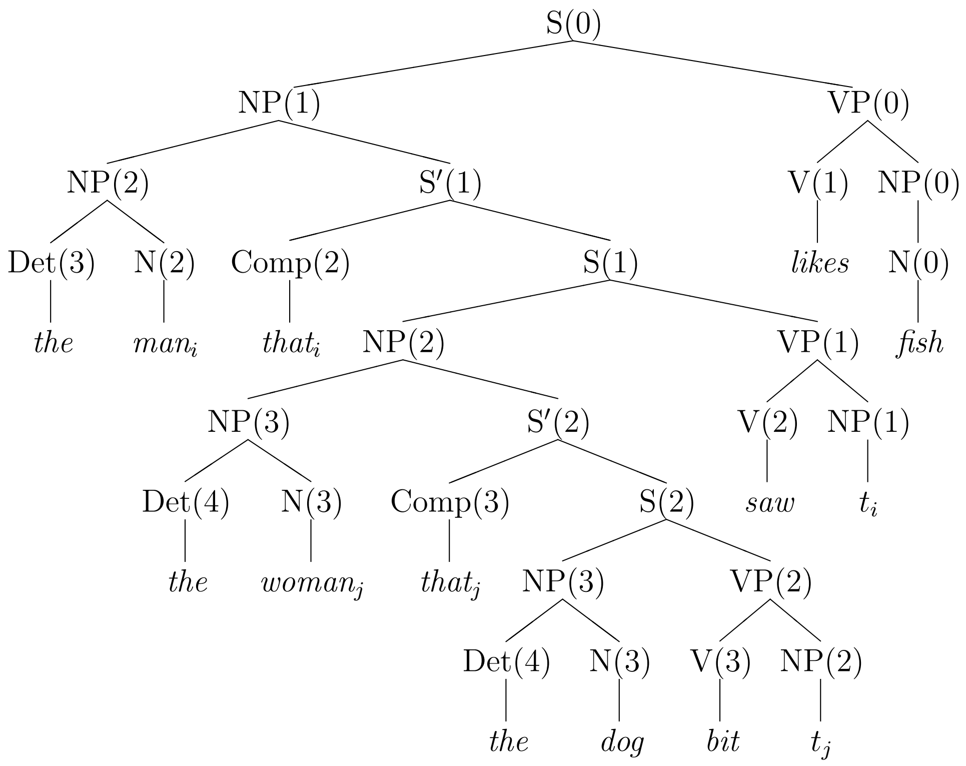


Figure 3: Yngve's top-down complexity measure for each of the nonterminals in the sentence *the man that the woman that the dog bit saw likes fish.*

The maximal complexity of (108) as measured by Yngve's metric occurs at the second and third Det nodes where four categories are waiting on the stack. At the second Det node the stack consists of an N node, an S′ node and two VP nodes. At the third Det node the stack consists of an N node and three VP nodes. Thus Yngve's hypothesis that the maximal memory load is somewhere around seven units appears to be too loose: according to this data, the borderline between acceptability and unacceptability occurs below four and above two of Yngve's memory units.

However, it turns out that it is not difficult to find sentences whose maximal complexity is three memory units which are acceptable and others whose maximal complexity is also three memory units that are unacceptable. Consider (109) and (110):

(109) The man the dog bit likes fish.

(110) # Men women dogs bite love like fish.

The maximal complexity associated with (109) occurs at the second occurrence of the category Det where there are three categories on the stack: an N and two VPs. The maximal complexity associated with (110) occurs at the third NP where there are three VP categories on the stack. Since (110) is completely unacceptable while (109) is completely acceptable, Yngve's complexity measurement makes an undesirable prediction.

While Yngve's complexity metric is close to making the right predictions with respect to the difference between right-branching and center-embedded structures, the metric is completely unworkable when it comes to left-branching structures (Miller & Chomsky (1963)). Predominantly left-branching languages like Japanese and Turkish are predicted by Yngve's complexity metric to be much harder to process than predominantly right-branching languages like English. However, there is no psychological evidence for such a conclusion. Center-embedded structures rather than right or left-branching structures are the source of processing difficulties in the world's languages. Consider (111) (from Kimball (1973)), an example of a left-branching structure from English, along with its structure and complexity measures in Figure 4:

(111) My cousin's aunt's dog's tail fell off.

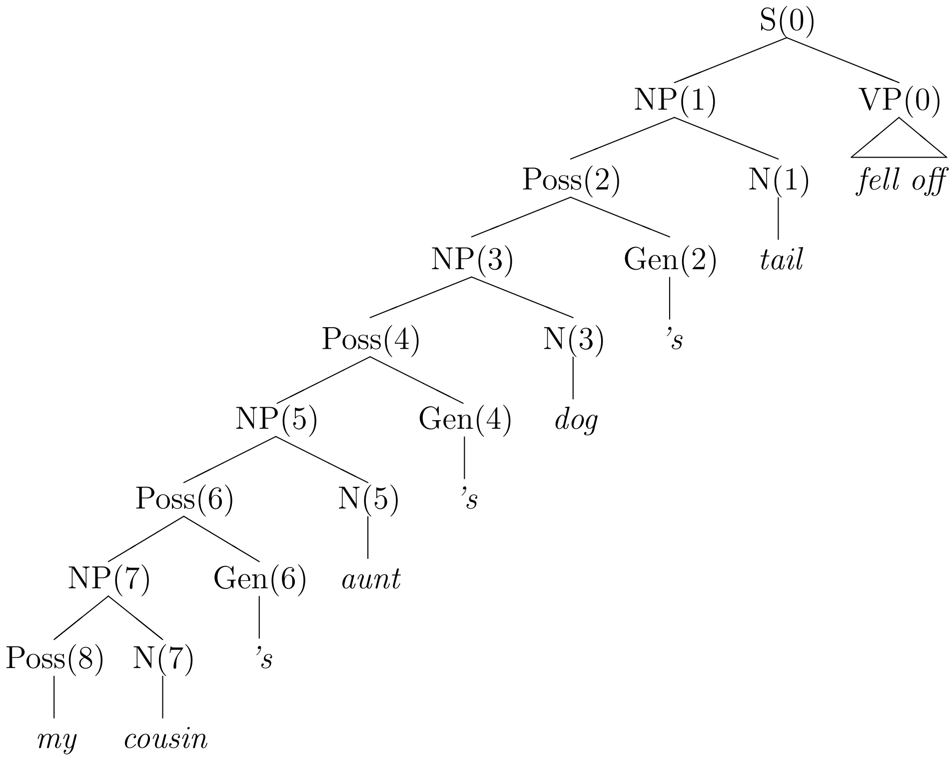


Figure 4: Yngve's top-down complexity measure for each of the nonterminals in the sentence *my cousin's aunt's dog's tail fell off.*

Sentence (111) is acceptable and yet its maximal complexity is eight of Yngve's memory units, far more than the maximal complexity associated with the unacceptable sentences (108) and (110). Thus Yngve's complexity metric makes the wrong predictions with respect to left-branching structures and is therefore not a psychologically plausible measure of human syntactic complexity.

### 3.2. Bottom-Up Parsing

The problem with Yngve's complexity measurement with respect to left-branching structures derives from the fact that the underlying algorithm is entirely top-down. While a completely bottom-up parsing algorithm would be able to process left-branching structures without the need for much stack space, a bottom-up algorithm predicts difficulty for right-branching structures (Chomsky & Miller (1963), Miller & Chomsky (1963)). The inadequacy of both the purely top-down and purely bottom-up parsing algorithms can be demonstrated by the use of a general parsing complexity metric which counts the total number of categories that need to be kept track of at any parse state by a particular parsing algorithm, whether this algorithm is top-down, bottom-up or a combination of the two. The categories that need to be stored at parse states can be placed in two groups: 1) the root categories of structures built thus far; and 2) the categories that are predicted somewhere below the current root categories but have not yet been found.

Yngve's metric is a top-down special case of this metric, in which categories which are predicted but not yet found make up the bulk of the complexity. In fact, Yngve does not include structures that have already been built in his complexity measure, although they must take up stack space under his assumptions. Yngve ignores these structures because he assumes a top-down generator: given top-down generation, there is always exactly one root structure, and so counting it does not affect the relative complexity of generating different structures. In processing with an entirely top-down algorithm there are at most two root categories to be considered at each parse state: the matrix category being predicted (*e.g.*, an S node), and the category of the current lexical item to be incorporated into the matrix structure. Thus the complexities given by Yngve's top-down metric for generation given in earlier figures need to be increased by an additional two units in most cases to give units of the more general complexity metric given above. Crucially, however, the observation that the complexity of left-branching structures increases without bound remains unchanged under the more general complexity metric.

In contrast to a top-down algorithm, all of the complexity in a bottom-up algorithm comes from storing completed categories which cannot yet be put together. Thus complexity increases without bound when considering right-branching structures as opposed to left-branching ones. As an illustration of this effect, complexity measures for each of the categories in the right-branching sentence in (107) are given in Figure 5 for a bottom-up parsing algorithm. For example, the bottom-up complexity measure associated with the first occurrence of the complementizer *that* is four units since four categories need to be stored at this point in order to get the structure in Figure 5: the NP *the dog*, the verb *saw*, the NP *the cat* and the complementizer *that*. The maximal bottom-up complexity associated with (107) is thirteen units, and in general, is unbounded on right-branching structures.

Since it is assumed that the number of categories that need to be kept track of in any parse

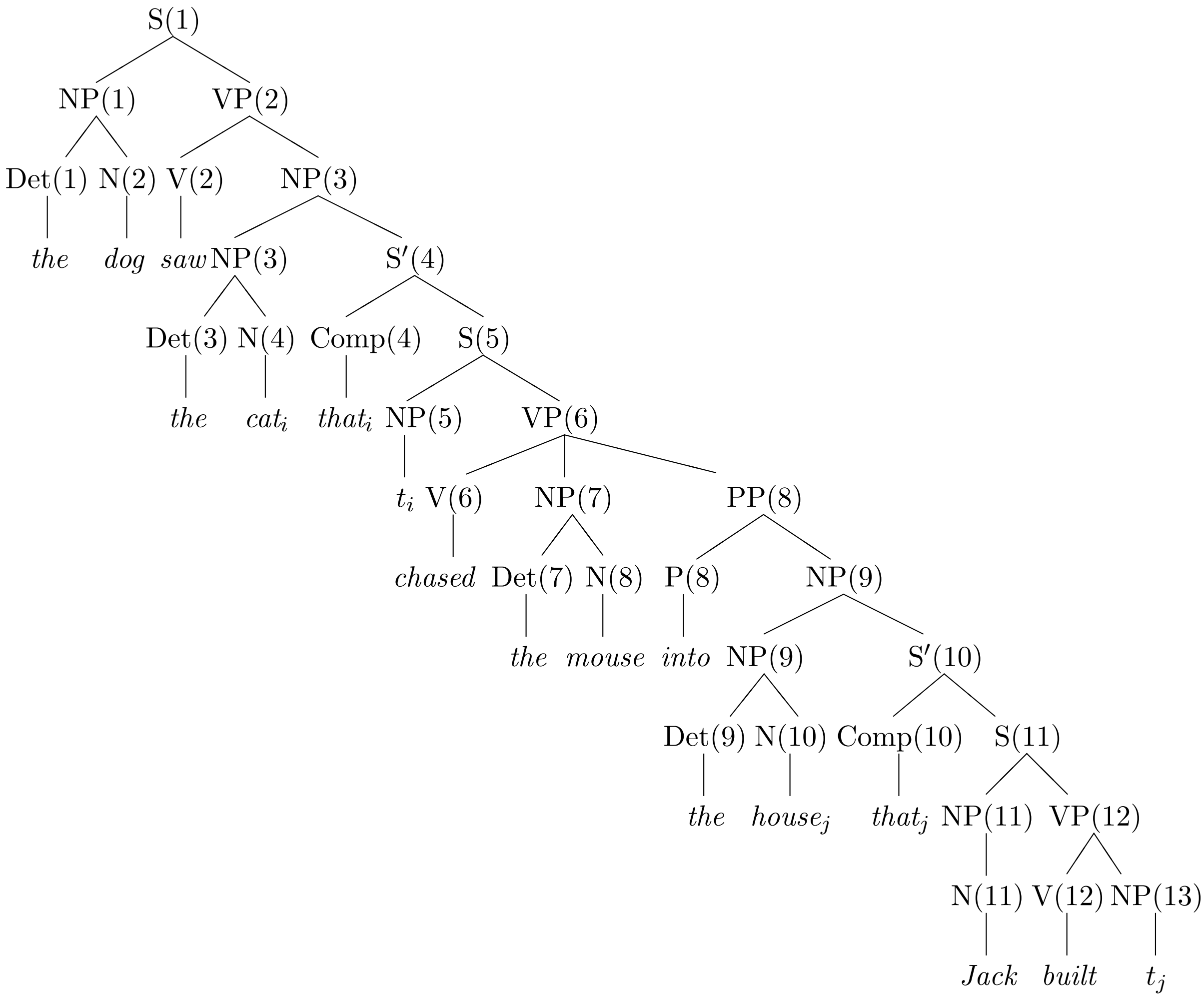


Figure 5: Maximal number of categories to be stored at the point of processing each nonterminal using a bottom-up parsing algorithm for the sentence *the dog saw the cat that chased the mouse into the house that Jack built.*

is bounded by the limited capacity of human short term memory, neither purely top-down nor purely bottom-up algorithms suffice as psychological parsing models.

### 3.3. Left-Corner Parsing

As a result of the problems with purely top-down and purely bottom-up parsing algorithms with respect to complexity as measured by stack space, a number of researchers have suggested that the human parser operates with both top-down and bottom-up components.[52] One such algorithm is a left-corner algorithm (see Aho & Ullman (1972) for a formal description of one such algorithm; see Kimball (1973, 1975), Frazier & Fodor (1978) and Johnson-Laird (1983) for left-corner variations for use as psychological models of human parsing). A left-corner parsing algorithm parses the leftmost category of the right hand side (RHS) of a grammar rule from the bottom up, and the rest of the grammar rule from the top down. Consider, for example, the following grammar rule:

(112) $L \rightarrow R_1 R_2 R_3$

The leftmost category on the RHS of the above rule is $R_1$. Hence this category is parsed from the bottom up. When a structure of category $R_1$ is built then a structure of category $L$ is proposed, along with categories $R_2$ and $R_3$ to the right. If the parser is successful in finding these categories, then category $L$ will be complete and it may cause left-corner prediction for all those rules in which $L$ is the leftmost category on the RHS. Consider then the left-to-right left-corner parse of sentence (114) given the following unambiguous context-free phrase structure grammar:

1. S → NP VP
2. VP → V NP
3. NP → Det N
4. Det → the
5. N → boy, man
6. V → saw

First, the word *the* is encountered. Since this item appears as the left-most element on the RHS of a rule, the LHS of this rule is built, the category Det, with the word *the* attached as its daughter. Next, since the category Det occurs as the leftmost category on the RHS of rule 3, an NP node is built, with the category Det attached below it to the left, and an unexpanded N node attached below it to the right (the top-down part of the algorithm). No further categories are built at this point since the NP node is not yet complete.

The word *man* is then input and the category N is built above it. This N then attaches as the rightmost daughter of the NP node which had just previously been constructed. At

[52]Chomsky & Miller (1963), Miller & Chomsky (1963) were among the first to make this suggestion, although they give no specific parsing algorithm.

this point an S node is built according to rule 1, with the now-complete left daughter and a projected VP attached below it as its right daughter.

The word *saw* is then input and the category V is built above it. This V attaches as the rightmost daughter of a projected VP, according to rule 2. This projected VP now attaches as the rightmost daughter of the previously projected S node, thus completing the S node.

The words *the boy* are parsed in the same way as the words *the man* were, and the resulting NP is attached as the rightmost daughter of the hypothesized VP, thus completing the parse.

Unlike purely top-down and purely bottom-up parsing algorithms, a left-corner parsing algorithm does not require excessive stack space for either left or right-branching structures. For example, the right-branching structure for sentence (107) is parsed with a left-corner algorithm while storing a maximum of six categories, as is depicted in Figure 6. Furthermore, the left-branching structure for (111) is obtained with a left-corner parser using at most three cells of stack space along the way, as is depicted in Figure 7. The difference in maximal complexities between the parses of these two structures derives form the fact that the predominantly right-branching structure in Figure 6 is not completely right-branching: it has a number of left-branching subcomponents. In fact, the complexity of a left-corner algorithm on strictly left or right binary branching structures is at most four units. Thus a left-corner parsing algorithm predicts no difficulty with either left or right-branching structures, as desired.

Furthermore, the metric that counts the number of categories that need to be kept track of at each parse state predicts increased complexity with center-embedded embedded structures. For example, the maximal complexity associated with the doubly center-embedded sentence in (108) is seven units, more than is necessary for the left or right-branching structures in Figures 6 and 7.

(108) # The man that the woman that the dog bit saw likes fish.

The complexities associated with the nonterminals in left-corner parsing of (108) are given in Figure 8.

While these results seem promising, it turns out that the maximal complexity of a left-corner parsing algorithm on a difficult sentence like (110) is only six units, the same as that associated with the right-branching structure in Figure 6.

(110) # Men women dogs bite love like fish.

Furthermore, the maximal complexity of (113) is seven units, despite the fact that this sentence is easy to process:

(113) The farmer chased the fox the dog bit into the woods.

Thus the number of categories locally stored by a left-corner parsing algorithm does not give a satisfactory metric for sentential complexity. This is not to say, however, that

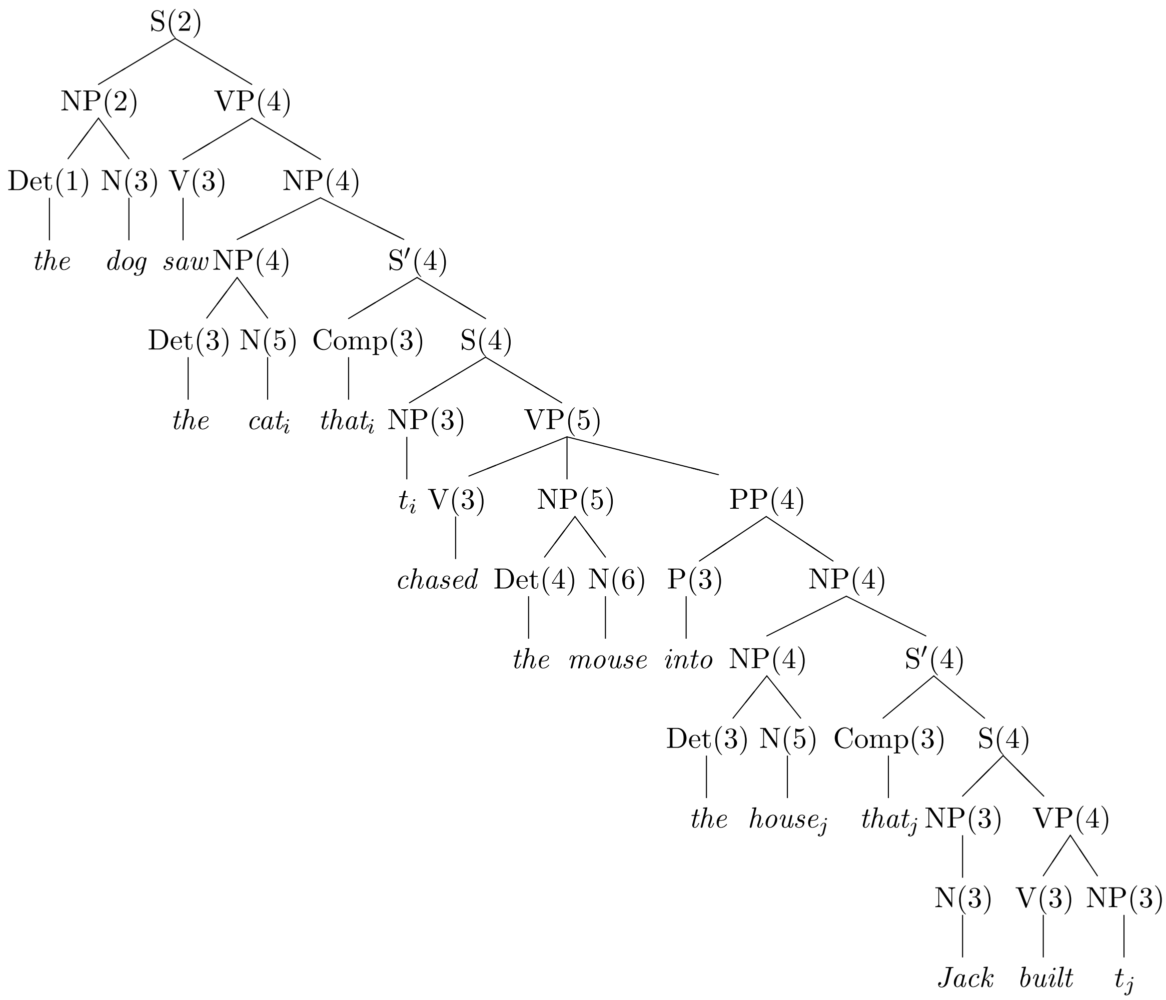


Figure 6: Maximal number of categories to be stored at the point of processing each nonterminal using a left-corner parsing algorithm for the sentence *the dog saw the cat that chased the mouse into the house that Jack built*.

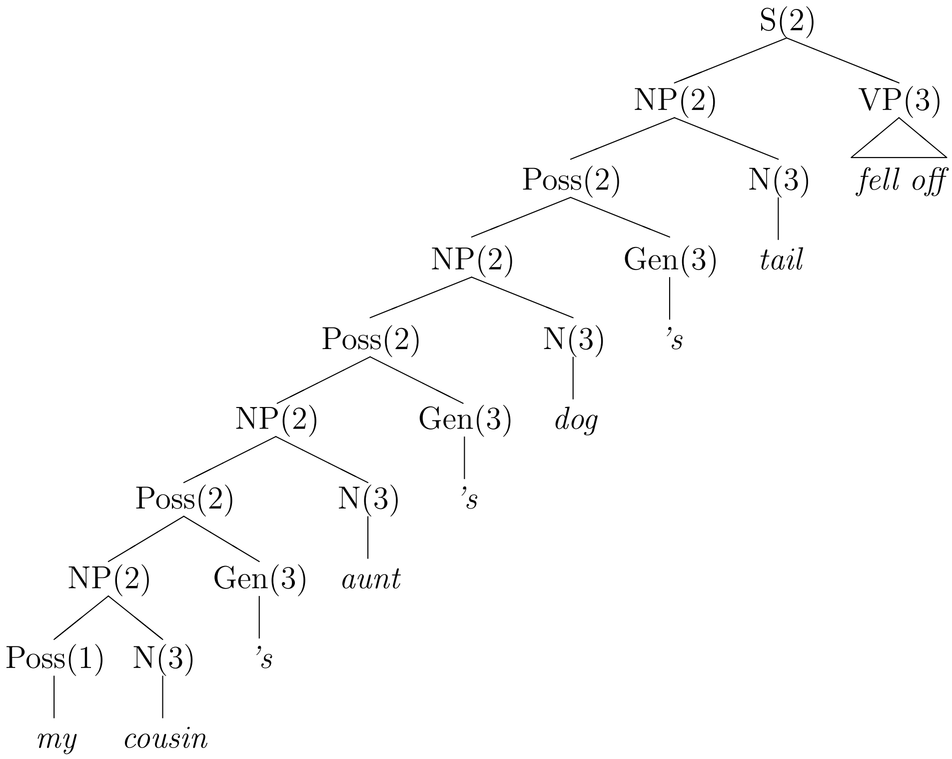


Figure 7: Maximal number of categories to be stored at the point of processing each non-terminal using a left-corner parsing algorithm for the sentence *my cousin's aunt's dog's tail fell off*.

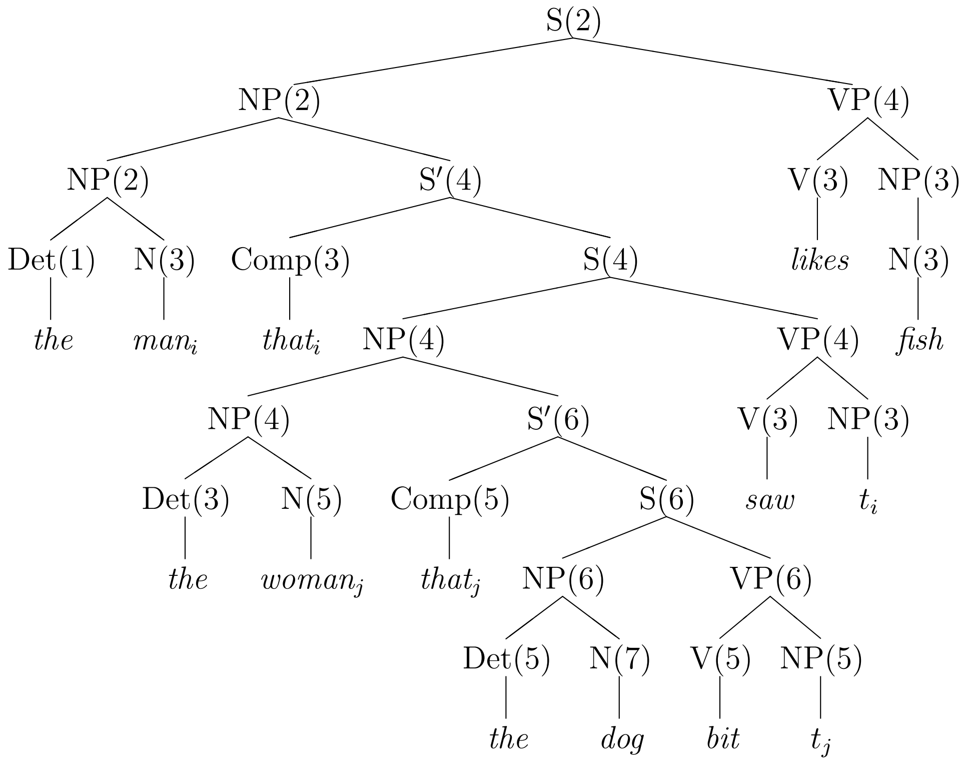


Figure 8: Maximal number of categories to be stored at the point of processing each non-terminal using a left-corner parsing algorithm for the sentence *the man that the woman that the dog bit saw likes fish*.

the human parser does not use a left-corner parsing algorithm. For instance, it could be that an appropriate metric for human sentential complexity is given by some function of the particular categories that are stored at a parse state. Thus while purely top-down and purely bottom-up parsing algorithms can be ruled out because of their unbounded local complexities with respect to left or right-branching structures, no such unbounded complexity occurs with a left-corner parsing algorithm.

### 3.4. Chomsky & Miller (1963)

As a result of the difficulty associated with purely top-down and purely bottom-up parsing algorithms, Chomsky & Miller (1963) and Miller & Chomsky (1963) assume that the human parsing algorithm has both top-down and bottom-up components, although they do not give a specific algorithm. Rather than trying to explain the difficulty associated with center-embedded structures in terms of a parsing algorithm, Chomsky & Miller proposed to account for this difficulty in two distinct ways. First Miller & Chomsky (1963) suggest that the ratio of terminal nodes to nonterminal nodes will make up an appropriate complexity metric. However, this approach is clearly unworkable, since it does not predict an asymmetry between center-embedded structures on the one hand and left and right-branching structures on the other.

Chomsky & Miller (1963) also suggested that an appropriate complexity metric might consist of the number of interruptions in grammatical relation assignment. Consider once again the center-embedded relative clause sentence in (108):

(108) # The man that the woman that the dog bit saw likes fish.

The verb *eats* needs to assign the grammatical relation SUBJECT to the NP *the man*, but this assignment is locally interrupted by an intermediate relative clause. Similarly, the verb *saw* needs to assign the grammatical relation SUBJECT to the NP *the woman*, but this assignment is also temporarily interrupted by another relative clause.

However, while Chomsky & Miller's hypothesis seems reasonable on the surface, it was not worked out well enough to make concrete predictions with respect to specific structures.[53]

### 3.5. Ratios of Words to Sentential Nodes: Fodor & Garrett (1967)

From Chomsky & Miller's suggestion that complexity be measured in terms of the ratio of terminal nodes to nonterminal nodes, Fodor & Garrett (1967) proposed that the perceptual complexity of a sentence could be measured by the ratio of terminal nodes (words) to sentential nodes in its associated structure, where the higher the ratio, the easier a sentence is to understand. Thus the complexity of (108) would be 12:3, while a simple declarative like (114) would be associated with a ratio of 5:1.

[53] In fact, Chomsky & Miller's insight underlies the theory of processing overload to be proposed in this thesis.

(114) The man saw the boy.

However, this metric of complexity is also problematic, since it is easy to construct acceptable sentences with ratios under 12:3. For example, the three word sentence *John saw Mary* has a ratio of three words to one sentence node, and thus Fodor & Garrett's metric wrongly predicts that it should be harder to understand than (108).[54]

### 3.6. The Principle of Two Sentences: Kimball (1973)

As a result of Chomsky & Miller's (1963) observations regarding the implausibility of either purely top-down or purely bottom-up parsing algorithms, Kimball (1973, 1975) proposes a hybrid algorithm, which he calls an over-the-top parsing algorithm.[55] This model is bottom-up in that structures are proposed based on the words that are input. But the algorithm is also partly top-down since predictions are made about what structures should appear next.

Given this parsing algorithm (see Kimball (1973) for more information on it), Kimball proposes a complexity metric consisting of the number of sentential nodes being considered at that point in the parse. Consider two classes of data. First consider the contrast between once and twice center-embedded relative clauses:

(115) The man that the woman likes eats fish.

(116) 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.

Second, consider the contrast between sentences with a simple sentential subject and sentences whose sentential subject contains a further sentential subject:

(117) a. That Joe left bothered Susan.
b. For John to smoke would annoy me.

(118) a. # That that Joe left bothered Susan surprised Max.
b. # That for Joe to leave bothers Susan surprised Max.
c. # That for John to smoke would annoy me is obvious.

Kimball (1973) accounts for the contrast in acceptability between each of these classes of

---

[54]Furthermore, a metric which measures local ratio is also unworkable by itself, since the local ratio associated with the formation of a single relative clause is no different from that with multiply embedded relative clauses.

[55]In fact, one of Kimball's reasons for abandoning bottom-up parsing algorithms was based on his misconception that bottom-up parsers cannot take account of left context. However, it possible for bottom-up parsers to account for left context. For example, a bottom-up shift-reduce parser takes account of left context by means of a suitable oracle (Pereira (1985)). In any case, Kimball's conclusion that the human parser should have both top-down and bottom-up components is still reasonable because of Chomsky & Miller's complexity observations.

sentences by appeal to the Principle of Two Sentences:

(119) The Principle of Two Sentences:
The constituents of no more than two sentences can be parsed at one time (Kimball (1973), p. 33).

The Principle of Two Sentences explains the unacceptability of sentences like those in (116) and (118), since, under Kimball's parsing assumptions, three incomplete sentence nodes are required in the parse of each of these sentences. Consider, for example, the parse of (116a) after the noun phrase *the dog* has been read:

(120) $[_S$ $[_{NP}$ the man $[_{S'}$ that $[_S$ $[_{NP}$ the woman $[_{S'}$ that $[_S$ $[_{NP}$ the dog ]]] ]]] ]]

At this point in the parse of (116a) there are three incomplete S nodes, so the sentence is ruled unacceptable by the Principle of Two Sentences. Since there is never a state in the parse of singly embedded relative clauses like (115) which requires more than two incomplete sentence nodes, the contrast between (115) and (116) is explained.

Similarly, (118c) violates this principle at the point of parsing the NP *John*:

(121) $[_S$ $[_{S'}$ that $[_S$ $[_{S'}$ for $[_S$ $[_{NP}$ John ]] ]] ]]

Structure (121) contains three incomplete sentence nodes, and thus (118c) is ruled unacceptable by the Principle of Two Sentences. Furthermore, since the sentences in (117) never require more than two incomplete sentence nodes, the contrast between (117) and (118) is explained.

Kimball's Principle of Two Sentences correctly predicts that two center-embedded relative clauses are unacceptable and having a sentential subject inside a sentential subject is unacceptable. However, it also predicts that having a center-embedded relative clause inside a sentential subject should be unacceptable. This is not the case, as is demonstrated by the examples in (122) ((122a) from Cowper (1976)):

(122) a. That the man who said that the book was finished didn't know his business is clear to everyone.
b. That the food that John ordered tasted good pleased him.
c. That the man that Mary loved died yesterday upset her.

The Principle of Two Sentences incorrectly predicts that these sentences should be unacceptable. At the point of processing the NP *John* in (122b), for example, there are three incomplete sentence nodes:

(123) $[_S$ $[_{S'}$ that $[_S$ $[_{NP}$ the food $[_{S'}$ that $[_S$ $[_{NP}$ John ]]] ]] ]]

Since there are three incomplete sentence nodes in (123), Kimball's Principle of Two Sentences wrongly predicts that (122b) is unacceptable.

Furthermore, Kimball's principle incorrectly predicts the following examples to be unacceptable ((124a) from Cowper (1976)):

(124) a. The possibility that the man who I hired is incompetent worries me.
b. The report that aliens who knew Elvis landed in Nashville was substantiated with photographs.

Each of these sentences contains a relative clause embedded inside a noun phrase complement. At the point of parsing the embedded relative clause there are three incomplete sentential nodes in each, and thus Kimball predicts these sentences to be unacceptable. However, these examples are perfectly acceptable, contrary to the predictions of the Principle of Two Sentences.

Additional evidence against Kimball's Principle of Two Sentences is provided by the following data:

(125) a. What the woman that John married likes is smoked salmon.
b. What the dog that bit Mary wanted was some dinner.

Each of these sentences contains a relative clause embedded inside a pseudo-cleft construction. At the point of parsing the subject of the relative clause, there are three incomplete sentential nodes in each and hence the Principle of Two Sentences is falsified. As a result of these difficulties, Kimball's principle cannot be correct as it stands.[56]

### 3.7. Sentential Subjects: Koster (1978)

An account proposed by Koster (1978) rules out sentences like those in (118) by appealing to the *grammar* rather than appealing to the processing complexity of the sentences. Consider once again (118c):

(118c) # That for John to smoke would annoy me is obvious.

Koster's analysis disallows the category $S'$ in subject position. In order to be interpreted as subject, Koster assumes that an $S'$ is base generated in topic position and then coindexed with a trace in subject position. The further assumption that only the root sentence can have a topic node gives the result that sentences like (118c) are ungrammatical. The $S'$ *for John to smoke* must be in a topic position to be licensed. However, there is no topic position in non-root sentences. Hence there is no position for this $S'$, and the sentence is ruled ungrammatical.

While Koster's (1978) account of multiple embedded sentential subjects deals satisfactorily with sentences like (118c), it also predicts that sentences like (126) are ungrammatical, an undesirable consequence:

[56] Kimball's Principle of Two Sentences has further problems since it relies heavily on another principle, the *Closure* principle. The Closure principle has many difficulties, some of which are outlined in Section 7.1.

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

Koster's analysis predicts that the sentential subject *for John to smoke* has to be in the specifier of S′, topicalized from the subject position. However, the head of this S′, *that*, occurs before the sentential subject, thus ruling out the topicalized position as a possible location. It turns out that there is no possible phrase structure for a sentence like (126) in Koster's analysis, so that this sentence is incorrectly predicted to be ungrammatical.

### 3.8. Frazier & Fodor (1978): the Sausage Machine

Frazier & Fodor (1978) propose an explanation of the difficulty people experience with center-embedded relative clauses in terms of a two stage parsing model called the Sausage Machine. The first stage of the model views the input string by means of a window which can see approximately six words at a time.[57] This stage assigns structure to the words that it sees in its window according to the rules in the grammar, and then passes these structures on to the second stage. The second stage of the parsing model then consults the grammar in order to put these small structures together. Frazier & Fodor claim that the difficulty that people experience with doubly center-embedded relative clauses can be explained by the two stages in the parsing model together with the limited viewing window of the first stage. However, this model is very difficult to evaluate since no explicit parsing algorithm is given for either of the two stages, nor is it made explicit exactly what causes the processing overload effect in center-embedded relative clause structures. In any case, consider Frazier & Fodor's description of the difficulty associated with (127) with respect to the Sausage Machine model:

(127) # The woman the man the girl loved met died.

Following Blumenthal (1966), Frazier & Fodor claim that the first six words of (127) are analyzed by the parser's first stage as a single conjoined noun phrase missing a conjunction. The next three words are similarly analyzed as a conjoined verb phrase. The second parsing stage then tries to put these structures together. Presumably the parser fails at some point after trying to put these structures together since the resulting structure is not a permissible representation for the input string, so reanalysis is necessary.[58] Eventually, the first stage

[57] Frazier & Fodor claim that the window width lies somewhere between five and seven words, but they give no formal definition as to what factors determine the window size. In order for their proposals to be testable, it is assumed, following Wanner (1980), that the window has a constant width of six words.

[58] Exactly how parses of doubly center-embedded relative clauses fail or succeed is not made explicit in Frazier & Fodor (1978). However, the difficulty experienced parsing (127) cannot be due only to the need for reanalysis during its processing, if indeed reanalysis is necessary. More generally, the explanation of processing overload effects cannot be identical to the explanation of garden-path effects in a psychologically plausible model. That is, while both effects present difficulty to the human parsing mechanism, it is possible to process garden-path sentences with relative ease once the correct structure for that sentence is known. The same is not true of a sentence which induces a processing overload effect: these kinds of sentences are still hard to interpret. Thus whatever explanation Frazier & Fodor give for processing overload sentences as in (127), the same explanation cannot be used for the difficulty with garden-path effects.

produces three separate NP structures for each of the NPs in (127) and three separate VP structures representing each of the verbs in (127). Frazier & Fodor's implicit claim is that putting all of these small constituents together is too big a job for the second stage of the parsing model.

Since lexical complementizers are included in (108), the difficulty with this sentence must derive completely from the consecutive verbs in its right perimeter under the Sausage Machine view:

(108) # The man that the woman that the dog bit saw likes fish.

As indicated earlier, it is difficult to evaluate the claims that Frazier & Fodor make regarding complexity, since the model that they propose is so vaguely presented that it cannot be properly tested. For example, it is not clear exactly what feature sentence (108) has that an appropriately modified acceptable alternative would not. Even so, there are clear empirical problems with the Sausage Machine treatment of complexity. First of all, if the difficulty associated with (127) and (108) derives from the fact that three verbs appear in rapid sequence after their stacked noun phrase subjects, then lengthening out the verb phrases so that each fits inside its own window should greatly reduce processing complexity. However, this does not seem to be the case, as is evidenced by (128):

(128) # The man that the woman that the dog bit on the leg saw at the large suburban mall likes fish.

(128) is still completely unacceptable, despite the fact that all of the structures that would be built by the first stage of the Sausage Machine model would be appropriate for use by the second stage. This result contradicts the predictions of the Frazier-Fodor model.[59]

Secondly, it was noted by Wanner (1980) that there exist doubly center-embedded sentences that consist of only six words. These sentences are uniformly difficult to process. Consider, for example, (130):

(130) # Women men girls love meet die.

Frazier & Fodor predict that a sentence like (130) should be easy to process since it can be analyzed entirely by the first stage processor. That is, all six words of the sentence can be seen at the same time and hence the grammar should assign these words the appropriate structure. This prediction is clearly wrong, since sentences like (130) are at least as hard as sentences like (127).

Not only do Frazier & Fodor predict that certain difficult sentences are easy, they also

[59]In fact, Frazier & Fodor attempt to support their model by appeal to examples similar to (128). For example, it is crucial for their model that (129) be acceptable:

(129) # The very beautiful young woman the man the girl loved met on a cruise ship in Maine died of cholera in 1962.

Following Milsark (1983), I find Frazier & Fodor's judgment of (129) to be acceptable (or very nearly acceptable) to be a dubious one. All of my informants agree that (129) is unacceptable.

predict that certain easy to process sentences are difficult. The model proposed by Frazier & Fodor predicts that any sentence which consists of a large number of very short constituents should cause give the parser a great deal of trouble (Milsark (1983). For example, consider (131) (from Milsark (1983)):

(131) I was amazed that Bill came, Susan left, John stayed, Mary ate, Jane drank...

The human parser should have as much trouble with (131) as it does with (127) and (108) according to the Sausage Machine model, but (131) gives the human parser no trouble, contrary to predictions.

As a result of the vagueness of the hypothesis together with its empirical shortcomings, Frazier & Fodor's model of linguistic complexity is rejected.

### 3.9. Frazier (1985)

Frazier (1985) hypothesizes that the complexity of a sentence need not be defined in terms of complexity at a single parse state: sentential complexity can also be defined in terms of complexity across a number of parse states. As a partial measure of complexity Frazier hypothesizes the *local nonterminal count* which is the sum of the value of all nonterminals introduced over any three adjacent terminals, where nonterminals are defined to be all nodes except for the immediate projections of lexical items. Furthermore, she states that each S or S′ node gets a value of 1.5 units, while all other nonterminals get a value of one unit. The measure to be computed for any given sentence is the *maximal local nonterminal count*: the largest such sum for a sentence.

Frazier assumes a partially top-down, partially bottom-up parsing algorithm somewhat like that of Kimball (1973). The nonterminal structures that she assumes to be built at each word, along with their associated complexities are given below for the parse of the sentence in (132):

(132) The men who sing love women.

(133)

| STRUCTURE | PROCESSING WEIGHT |
|---|---|
| $[_S\ [_{NP}$ the ]] | 2.5 |
| men | 0 |
| $[_{NP}\ [_{S'}$ who ]] | 2.5 |
| $[_S\ [_{VP}$ sing ]] | 2.5 |
| $[_{VP}$ love ] | 1 |
| $[_{NP}$ women ] | 1 |

Thus the maximal local nonterminal count associated with the parse of (132) is 6 units, which is the nonterminal count associated with the words *who sing love*.

Frazier's claim regarding sentential complexity is as follows: given two sentences with an equal overall nonterminal ratio, the sentence with the higher maximal local nonterminal count will be more difficult to process.[60] Evidence that she gives in favor of this claim is given in (134):

(134) a. That men who were appointed didn't bother the liberals wasn't remarked upon by the press.

b. # That that men were appointed didn't bother the liberals wasn't remarked upon by the press.

c. # That that that men were appointed didn't bother the liberals wasn't remarked upon by the press upset many women.

Sentence (134a), which contains a relative clause embedded inside a sentential subject is acceptable, while sentences (134b) and (134c), which both include multiply embedded sentential subjects, are not. Furthermore, (134c) is less acceptable that (134b), owing to the fact that it has a further degree of sentential subject embedding than (134b). According to Frazier, these relative judgments may be explained by appeal to the maximal local nonterminal count in each. The maximal local nonterminal counts for the parses of the sentences in (134) are 8, 8.5 and 9 units respectively. In each case this maximal count occurs with respect to the first three words in each sentence. The nonterminal structures along with their associated complexities are given below for the parse of the first three words of each of the sentences in (134):

(135)

| STRUCTURE | PROCESSING WEIGHT |
|---|---|
| $[_S\ [_{S'}$ that ]] | 3 |
| $[_S\ [_{NP}$ men ]] | 2.5 |
| $[_{NP}\ [_{S'}$ who ]] | 2.5 |
| | |
| $[_S\ [_{S'}$ that ]] | 3 |
| $[_S\ [_{S'}$ that ]] | 3 |
| $[_S\ [_{NP}$ men ]] | 2.5 |
| | |
| $[_S\ [_{S'}$ that ]] | 3 |
| $[_S\ [_{S'}$ that ]] | 3 |
| $[_S\ [_{S'}$ that ]] | 3 |

Since the maximal local nonterminal count is least in (134a), slightly higher in (134b) and highest in (134c), the relative complexities are explained. That is, presumably the maximum possible local nonterminal count is eight units, thus explaining the fact that (134a) is just acceptable while the others are worse.

[60] Although Frazier claims that only sentences with the same overall ratio of nonterminal to words can be compared via the maximal local nonterminal count, much of her evidence consists of comparisons between sentences that do not have exactly the same nonterminal to word ratios, although the ratios are always close.

Furthermore, Frazier claims that her system explains the difference in processing difficulty between (136) and (137):

(136)
That Ray smiled pleased Sue.
3 2.5 1 1 1

(137)
It pleased Sue that Ray smiled.
2.5 1 1 1.5 2.5 1

The numbers below each word indicate the nonterminal count associates with that word in the parse. The maximal nonterminal count associated with (136) is 6.5 units, for the first three words of this sentence *that Ray smiled*. On the other hand, the maximal nonterminal count associated with (137) is only 5 units, for the last three words of this sentence *that Ray smiled*. Since (137) is easier to process than (136), Frazier's complexity metric makes the appropriate prediction.

Finally, Frazier claims that her complexity metric correctly predicts the difference between singly and doubly center-embedded relative clauses as well as the difference between having center-embedded relative clauses in subject and object positions (data from Eady & Fodor (1981)):

(138)
The patient the nurse sent to the doctor the clinic had hired met Jack.
2.5 0 5 0 1 2 1 0 5 0 1 0 2 1

(139)
?# Jack met the patient the nurse the clinic had hired sent to the doctor.
2.5 1 1 0 5 0 5 0 1 0 2 2 2 1

(140)
# The patient the nurse the clinic had hired sent to the doctor met Jack.
2.5 0 5 0 5 0 1 0 2 2 1 0 1 1

Eady & Fodor (1981) found that sentences like (138), which contains only singly center-embedded relative clauses, were significantly easier to process than sentences like (139), which contained a doubly center-embedded relative clause. Furthermore, Eady & Fodor found that sentences in which the doubly center-embedded relative clause were in object position, as in (139), were significantly easier to process than corresponding sentences in which the doubly center-embedded relative clause was in subject position as in (140). Frazier claims that her metric correctly predicts these facts in the following way. First, the maximal local nonterminal count associated with (138) is 7.5 units, over the first three words *the patient the*. The maximal nonterminal count associated with both sentences (139) and (140) is 10 units, over the words *the nurse the* in both examples, thus partially explaining the difference in complexity between singly and doubly center-embedded relative clauses. Furthermore, Frazier claims that the explanation for the difference in complexities between (139) and (140) is due to a difference in the second highest local nonterminal count in these two sentences. The second highest local nonterminal count in (140) is 7.5 units (over the first three words *the patient the*), while the second highest local nonterminal count in (139) is only 6 units, occurring in three different locations in (139).

While Frazier's approach seems plausible given the data that she examines, it turns out that there are serious problems with her metric of complexity calculation: it does not work for center-embedded relative clauses as she implicitly claims. Consider the strong contrast between (141) and (142):

(141)
The man that the dog bit ate the cake that the woman saw.
2.5 0 2.5 2.5 0 2 1 1 0 2.5 2.5 0 2

(142)
# The man that the woman that the dog bit saw ate the cake.
2.5 0 2.5 2.5 0 2.5 2.5 0 2 2 1 1 0

(141) contains only singly embedded relative clauses and is acceptable. (142), on the other hand, contains a doubly center-embedded relative clause in subject position and is therefore unacceptable. However, Frazier's metric does not distinguish the complexities of these two sentences. Each of the two sentences has the same number of words, the same number of nonterminals and the same local nonterminal count profile. In particular, the maximal local nonterminal count in each of these two sentences is 5 units, much lower than the maximal local nonterminal counts of other acceptable and unacceptable sentences (see *e.g.* (136) and (134b)).

For Frazier's approach to predict unacceptability, it must be the case that three words with high complexity be immediately adjacent. This is just not always the case in sentences that are difficult to comprehend. Center-embedded relative clauses become unacceptable when one relative clause is embedded inside another, resulting in three adjacent noun phrases in the input string. For Frazier's metric to work, three words must encompass the initial word of the second NP and the initial word of the third NP. Since 1) a complementizer may intervene and 2) noun phrases may consist of arbitrarily many words, there is no guarantee that Frazier's window of three words will encompass the desired words. In fact, no constant length window will suffice to predict unacceptability and variations in complexity.

Thus the only doubly center-embedded sentences that Frazier's complexity metric correctly predicts to be difficult are those in which the most deeply embedded relative clause has no lexical complementizer and in which the second NP consists of two words or less. Hence Frazier's theory of center-embedded relative clauses is unsatisfactory as it stands. The only other data that her theory accounts for consists of sentential subject data, and even so, her theory crucially relies on the stipulation that the nonterminal nodes S and S′ receive a load of 1.5 units while other nonterminals receive a load of one unit: no explanation or independent evidence is given for this assumption. As a result of these serious problems, Frazier's proposal regarding complexity must be rejected.

# 4. The Underlying Parser

This chapter describes the underlying parsing model to which the constraints outlined in Chapter 1 are applied. Section 4.1 gives an overview of some of the motivations for a partly bottom-up, partly top-down parsing algorithm along with a brief introduction to the left-corner parsing algorithm assumed here. The data structures utilized by this algorithm are given in Section 4.2 and the underlying grammatical and lexical representation assumptions are given in Section 4.3. Section 4.4 then gives the full parsing algorithm along with example parses.

## 4.1. Left-Corner Parsing Revisited

The work of a number of psycholinguistic researchers has established that linguistic input is processed incrementally (see Marslen-Wilson (1975, 1987), Tyler & Marslen-Wilson (1977), Swinney (1979), Marslen-Wilson & Tyler (1980), Shillcock (1982), Garrod & Sanford (1985), and Tanenhaus, Carlson & Seidenberg (1985) among others). Thus any model of human linguistic processing should construct representations in such a way so that incomplete input is associated with structure. Furthermore, as noted in Section 3.4, Chomsky & Miller (1963) and Miller & Chomsky (1963) observed that neither purely top-down nor purely bottom-up parsing algorithms would suffice as models of human linguistic processing since each algorithm predicts difficulty with either left or right-branching structures, structures that people have no difficulty with. As a result, the human parser must have both top-down and bottom-up components.

As a result of the desire for a parsing algorithm that 1) associates incomplete input with structure and 2) has both top-down and bottom-up components, it is proposed, following Kimball (1975) and Johnson-Laird (1983) among others, that the human parser operates using a left-corner parsing algorithm (Aho & Ullman (1972)). Recall from Section 3.3 that a left-corner parsing algorithm parses the leftmost category of the right hand side (RHS) of each grammar rule from the bottom-up and the rest of the grammar rule from the top down.

(143) $L \rightarrow R_1 R_2 R_3$

Thus when a structure of category $R_1$ is found, a structure of category $L$ is also built above $R_1$, predicting the occurrence of categories $R_2$ and $R_3$ to the right.

The left-corner parsing algorithm that I will assume differs from the one discussed in Section 3.3 in that it driven by linguistic heads (*cf.* Proudian & Pollard (1985)). The algorithm described in Section 3.3 allows the hypothesis of a left-hand side category $L$ only when a *complete* structure of left corner category $R_1$ was found. The left-corner algorithm that I propose allows the hypothesis of category $L$ whenever $R_1$'s head category has been located, where the *head* of a structure is determined by $\overline{\text{X}}$ Theory (Jackendoff (1977), Chomsky (1986b); see also Pollard (1984)). The complete parsing algorithm is given in Section 4.4.

## 4.2. Data Structures

A number of psycholinguistic studies performed in the late 1970s and early 1980s arrived at the conclusion that people access multiple readings for a word in parallel independent of context (see Swinney (1979), Tanenhaus, Leiman & Seidenberg (1979), and Seidenberg, Waters, Sanders & Langer (1984) among others). In order to account for this result, any model of human linguistic processing must also access all lexical entries for a word, independent of context. Furthermore, as outlined in Chapter 1, it is hypothesized that the human parser allows multiple representations for an input string (see Crain & Steedman (1985), Kurtzman (1985), Gorrell (1987), Schubert (1984, 1986), Waltz & Pollack (1985), Taraban & McClelland (1988), and Trueswell *et al* (1989) among others). Thus it is proposed that the parallel lexical access extends beyond the lexicon.

As a result of the desire to maintain multiple representations for an input string, the data structures employed by the parsing model proposed here allow no limit in principle to the number of representations for the input string nor do they limit the way in which the input string is partitioned. Rather, it is the job of the constraints whose general format is outlined in Chapter 1 to limit the number and kind of representations for the input string. The data structures upon which the parsing model rests are called the *buffer* and the *stack-set* (*cf.* Marcus (1980), Gibson (1987), Gibson & Clark (1987), and Clark & Gibson (1988)). When a word is input to the proposed parsing 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 *stack-set* contains a set of stacks of buffer cells each of which contains tree structures for previous input. The parser builds trees in parallel based on possible attachments between the buffer and the top of each stack in the stack-set. The buffer and stack-set are formally defined in (144) and (145).

(144) 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 at most one buffer cell.

(145) The *stack-set* is a set of stacks of buffer cells, where each stack in this set represents the same segment of the input string as every other stack in this set.

$$\begin{array}{l} \{ ( \{ S_{1,1,1}, S_{1,1,2}, ..., S_{1,1,n_{1,1}} \}, \{ S_{1,2,1}, S_{1,2,2}, ..., S_{1,2,n_{1,2}} \}, ... \{ S_{1,m_1,1}, S_{1,m_1,2}, ..., S_{1,m_1,n_{1,m_1}} \} ) \\ \quad ( \{ S_{2,1,1}, S_{2,1,2}, ..., S_{2,1,n_{2,1}} \}, \{ S_{2,2,1}, S_{2,2,2}, ..., S_{2,2,n_{2,2}} \}, ... \{ S_{2,m_2,1}, S_{2,m_2,2}, ..., S_{2,m_2,n_{2,m_2}} \} ) \\ \quad ... \\ \quad ( \{ S_{p,1,1}, S_{p,1,2}, ..., S_{p,1,n_{p,1}} \}, \{ S_{p,2,1}, S_{p,2,2}, ..., S_{p,2,n_{p,2}} \}, ... \{ S_{p,m_p,1}, S_{p,m_p,2}, ..., S_{p,m_p,n_{p,m_p}} \} ) \} \end{array}$$

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 $j$th buffer cell of stack $i$.

That is, given a stack $\sigma_i = (\tau_1, ..., \tau_{m_i})$, where each $\tau_j$ is a set of tree structures, then for each $j = 1, ..., m_i$, every tree structure in $\tau_j$ represents the same segment of the input string,

$\phi_{\sigma_i,j}$. Furthermore, each stack $\sigma_i$ in the stack-set represents a certain ordered partition $\pi_i$ of the current prefix of the input string where $\pi_i =< \phi_{\sigma_i,1}, ..., \phi_{\sigma_i,m_i} >$.

The motivation for the stack-set and buffer 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, where there exist only two parser operations: attachment of one structure to another 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. For example, consider a parser state in which the buffer is non-empty and the stack-set contains only a single cell stack:

(146)

Stack-set: $\{ ( \{ S_1, S_2, ..., S_n \} ) \}$
Buffer: $\{ B_1, B_2, ..., B_m \}$

Suppose that attachments are possible between the buffer and the single stack cell. The structures that result from these attachments will take up a single stack cell. Let us call these resultant structures $A_1, A_2, ..., A_k$. If all possible operations are to take place at this parser state, then the contents of the current buffer must also be pushed on top of the current stack. Thus two stacks, both representing the same segment of the input string, will make up the resulting stack-set:

(147)

Stack-set: $\{ ( \{ A_1, A_2, ..., A_k \} )$
$( \{ B_1, B_2, ..., B_m \} \{ S_1, S_2, ..., S_n \} ) \}$
Buffer: $\{ \}$

Since these two stacks break up the same segment of the input string in different ways, the stack-set data structure is necessary. An example parse using the unconstrained parsing algorithm is given in Section 4.4.5. Before this example parse can be given, it is necessary to describe what structures are built based on the lexicon and grammar and how these structures are attached to one another.

## 4.3. The Grammar and Lexicon

The grammar and lexical representations that I assume are based on current work in a number of syntactic theories, most notably Government-Binding theory (see *e.g.* Chomsky (1981, 1986a, 1986b)), but also Head-driven Phrase Structure Grammar (HPSG) (Pollard & Sag (1987)) and Lexical Functional Grammar (LFG) (Kaplan & Bresnan (1982)). The assumptions regarding phrase structure are given in Section 4.3.1. Example rules are also given in this section. The structure of the relevant components of a lexical entry are given in Section 4.3.4.

### 4.3.1. $\overline{\text{X}}$ Theory

I assume a version of $\overline{\text{X}}$ Theory very like that which is presented in Chomsky (1986b) (see also Jackendoff (1977), Pollard (1984)). The theory that I assume posits the existence of the five following schematic rules, where the category Y is a variable indicating any syntactic category, and the order of the right hand sides is dependent upon language, category and lexical item:

(148) a. $Y'' \rightarrow$ (Spec,) $Y'$
b. $Y' \rightarrow Arg^*$, Y
c. $Y'' \rightarrow$ Mod, $Y''$
d. $Y' \rightarrow$ Mod, $Y'$
e. $Y'' \rightarrow$ Mod, $Y''$

The categories Y, $Y'$ and $Y''$ in the above schematic rules are all assumed to be coindexed, so that all features of any one are identical to the features of the others. The category Y (sometimes called Y0) is referred to as the *zero-level projection* of Y or the *minimal projection* of Y. The category $Y'$ ($\overline{\text{Y}}$) is called the *one-bar projection* of Y. The category $Y''$ ($\overline{\overline{\text{Y}}}$, YP) is called the *two-bar projection* of Y.[61] The highest double bar projection of a category is called the *maximal projection* of that category. The *head* of an $\overline{\text{X}}$ category is its zero-level projection. $\overline{\text{X}}$ Theory also includes the assumption that every structure has a head.[62,63] Thus the presence of the category $Y''$ implies the existence of categories $Y'$ and Y below it.

In English, specifiers (*e.g.*, subjects) appear to the left of their heads, while arguments (complements) appear to the right. Modifiers can appear on either the left or the right of a head, depending on the categories involved. A sample English $\overline{\text{X}}$ structure is given in Figure 9.

### 4.3.2. Rules vs. Lexical Requirements

Recently it has been claimed that all phrase structure rules can be purged from the grammar to be replaced by lexical requirement specifications and filters upon representations (see *e.g*,

[61] Some versions of $\overline{\text{X}}$ Theory assume three bar levels of projections.

[62] In the GB literature it is commonly assumed that the five schemata in (148) follow from the interaction of other syntactic principles with the assumption (principle) that all structures have heads (see *e.g.*, Stowell (1981)). However, it is not currently clear how these derivations take place. Thus I will assume the schemata in (148) as basic to the theory.

In any case, it is probably desirable to implement $\overline{\text{X}}$ Theory as the schemata in (148) rather than some underlying principles for parsing efficiency reasons. In particular, it would not be desirable to have to derive the schemata in (148) every time a sentence is parsed.

[63] It is also commonly assumed that all structures have two-bar projections, an assumption that I do not make here.

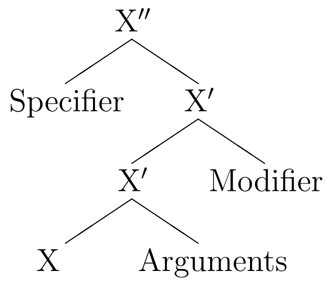


Figure 9: Example English $\overline{\mathrm{X}}$ structure

Stowell (1981), Chomsky (1986a)).[64] For example, it is thought that the rule NP → Det N′ can be replaced by the requirement that the category Det, if it exists in a given language, always acts as specifier for the category NP, where the location of the specifier with respect to the head of the NP is determined by general $\overline{\mathrm{X}}$ principles along with the word order properties of the language being considered. Thus it is a lexical property of the category Det that determines its relationship to other categories. Similarly, the fact that a PP can modify an NP is a lexical property of nouns and prepositions.

While the position that phrase structure rules can be replaced by more general principles makes sense with respect to the theory of grammar, note that it need not apply to processing theory. That is, it may be that the grammar contains no phrase structure rules, but rather, consists of more general principles which derive the rules. Even if this is the case, the device which uses such knowledge may use it in a compiled form: some phrase structure rules, for example. Thus while the knowledge of language might not contain any phrase structure rules, the linguistic processor might still use such rules. Hence my assumption of the existence of some kinds of pre-compiled linguistic knowledge for use by the processor does not contradict current assumptions regarding the structure of the grammar.

### 4.3.3. Sample $\overline{\mathrm{X}}$ Rules

Following Chomsky (1986b), I assume that all categories conform to $\overline{\mathrm{X}}$ Theory. In particular, the traditional categories S and S′ are assumed to be replaced by the categories Infl (I) and Comp (C) respectively, each of which has a two-bar maximal projection.[65] These categories are known as non-lexical categories since they need not have any lexical content. The category Infl contains tense and agreement information which may or may not be lexically realized. Infl takes a verb phrase as its complement and obligatorily requires an NP or CP specifier by the Extended Projection Principle (Chomsky (1981, 1982)). The category Comp takes an Infl phrase as its argument and allows a major category specifier, where a major category is one from the set { Comp, Infl, Noun, Verb, Adjective, Preposition, Adverb }. The behavior of these two categories can be summarized by the following $\overline{\mathrm{X}}$ rules:

[64] See also Berwick (1987) and Abney (1988) for comments on what counts as a "principle-based" parser as opposed to a rule-based one.

[65] See Pollock (1989) for the description of a theory which divides the category Infl into further categories.

(149) a. IP → {NP, CP} I′
b. I′ → I VP
c. CP → XP C′, where X is a major category
d. C′ → C IP

All argument and some specifier $\overline{\text{X}}$ rules need not be given in the grammar for lexical categories, since these rules are derivable from the lexical entries for these categories. See Section 4.3.4 for a description of the relevant contents of a lexical entry. However, all modifier and some specifier rules which apply across classes of lexical items are assumed to be present in the grammar (extended lexicon). In particular, I assume the existence of $\overline{\text{X}}$ rules like the following:

(150) a. NP → DetP N′
b. $N'_1 \rightarrow N\ N'_2$
c. $N'_1 \rightarrow N'_2$ PP
d. $N'_1 \rightarrow N'_2$ CP

Thus a determiner phrase (DetP) may be the specifier of a noun phrase by rule (150a). A zero-level projection of a noun may appear as a pre-head modifier of a one-bar projection of a noun by rule (150b). This rule allows English nominal compounds. Rules (150c) and (150d) allow PPs and relative clause CPs to modify noun phrases at the one-bar projection level. See Gibson (in preparation) for a listing of the grammatical rules used in the current implementation.

### 4.3.4. Lexical Entries

I assume that lexical entries contain the following components:

- Lexical string – the word in question;
- Syntactic category;
- Agreement features (*e.g.*, person, number, gender, case, tense);
- A $\theta$-grid.

A *$\theta$-grid* consists of a ordered list of *$\theta$-structures* dictating the subcategorization pattern of given lexical item (Pollard & Sag (1987), *cf.* Stowell (1981)). A $\theta$-structure is assumed to contain the following components:

- The subcategorized category. *e.g.*, NP, CP, VP, AP, PP, IP.
- Thematic role. *e.g.*, AGENT, THEME, GOAL, PROPOSITION, *etc.* Although there is usually some correlation between the thematic role and the subcategorized category

in a $\theta$-structure (Grimshaw (1981), Pesetsky (1983), Chomsky (1986a)), I assume that the subcategorized category is not completely determined by the thematic role, so that both must be explicitly represented in the $\theta$-structure (see *e.g.*, Pollard & Sag (1987)).

This component of a $\theta$-structure may also be left out if no thematic role is assigned in the subcategorization. For example, the complementizer *that* assigns no thematic role to its subcategorized Infl phrase and hence no role is included the corresponding $\theta$-structure.

- A value for the binary feature *external*, which indicates whether the subcategorized category is assigned inside the $\overline{\text{X}}$ single bar projection (external = NIL) or to the specifier (subject) position of the maximal projection of the head in question (external = T). At most one $\theta$-structure in a $\theta$-grid is marked as external.

- A value for the binary feature *oblig*, which indicates whether the subcategorized category is obligatory in the syntax.[66]

Furthermore, it is assumed that the order of the thematic role receiving categories with respect to the head in English is determined by the order of internal (external = NIL) $\theta$-structures inside a $\theta$-grid, such that earlier internal arguments in a $\theta$-grid must appear closer to the head than subsequent internal arguments (Pollard & Sag (1987)).[67]

The thematic role component of a $\theta$-structure may be empty if the lexical entry in question is a *functional* element as opposed to thematic element. Following recent 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). There are a number of properties that distinguish functional elements from thematic elements, the most crucial being that functional elements mark grammatical or relational features while thematic elements pick out a class of objects or events. Thematic categories include nouns, verbs, adjectives and some prepositions; functional categories include determiners, complementizers, inflection markers and other prepositions. Since a complementizer is a functional element, it subcategorizes for an IP, but assigns no thematic role to that IP. Similarly, the category Infl subcategorizes for a VP, but assigns no thematic role to that VP.

In addition to the components given above, I also assume the existence of semantic, pragmatic and discourse information on lexical entries, although I give only syntactic information here. I leave the definition of more complete lexical representations to future work.

---

[66]Determining which phrases count as arguments and which count as adjuncts is not always easy. See Pollard & Sag (1987) for a list of some syntactic tests that can be used in making this determination.

[67]The assumption that a $\theta$-grid is ordered is made in order to account for English word order phenomena. In Chomsky (1981), Stowell (1981) and in subsequent works, it is assumed that the theory abstract Case assignment explains English word order. As long as the empirical predictions are correct, Case theory could replace the $\theta$-grid ordering assumption given here. No empirical consequences rest on these assumptions with respect to the processing theory presented in this thesis.

Examples will make the definition of a $\theta$-grid more clear. First consider the following $\theta$-grid for a functional category, the complementizer *that*:

(151)

( ( Subcat = IP, Oblig = T, External = NIL ) )

This $\theta$-grid indicates that the complementizer *that* requires the syntactically obligatory IP complement, to which it assigns no thematic role.

Consider now the verb *send*, which I assume has two $\theta$-grids, the first of which is given in (152) (*cf.* Kayne (1984), Larson (1988) and the references in each):

(152)

( ( Subcat = NP, Thematic-Role = AGENT, Oblig = T, External = T )
( Subcat = NP, Thematic-Role = THEME, Oblig = T, External = NIL )
( Subcat = PP, Thematic-Role = GOAL, Oblig = NIL, External = NIL ) )

The above $\theta$-grid is initiated by an external NP agent $\theta$-structure indicating that this $\theta$-structure is to be associated with the subject of the VP headed by *send*. The other two arguments in this $\theta$-grid are the internal arguments of one reading of the verb *send*. Since the theme NP occurs before the goal PP in the $\theta$-grid, the theme NP prefers to be closer to the head than the goal PP, as is evidenced by the examples in (153):

(153) a. John sent the flowers to Mary.
b. ? John sent to Mary the flowers.

Furthermore, the goal PP $\theta$-structure is marked as optional because of the grammaticality of sentences like (154):

(154) John sent the flowers.

Consider now a second $\theta$-grid for the verb *send*:

(155)

( ( Subcat = NP, Thematic-Role = AGENT, Oblig = T, External = T )
( Subcat = NP, Thematic-Role = GOAL, Oblig = T, External = NIL )
( Subcat = NP, Thematic-Role = THEME, Oblig = T, External = NIL ) )

The $\theta$-grid in (155) is the so-called dative-shifted version of the $\theta$-grid in (152). As in the $\theta$-grid in (152), the $\theta$-grid in (155) contains three $\theta$-structures, one of which is marked external, the NP agent. In contrast to (152), however, both of the internal arguments of (155) are obligatory. Thus (156) cannot be parsed using this $\theta$-grid:

(156) John sent Mary.

Furthermore, the order of the two internal arguments in (155) is reversed with respect to (152). Thus the only parse of (157) links the NP *Mary* to the $\theta$-role GOAL and the NP *the flowers* to the NP THEME:

(157) John sent Mary the flowers.

Consider now the following $\theta$-grid for one lexical entry of the verb *speak*:

(158)

( ( Subcat = NP, Thematic-Role = AGENT, Oblig = T, External = T )
( Subcat = PP, Thematic-Role = GOAL, Oblig = NIL, External = NIL )
( Subcat = PP, Thematic-Role = THEME, Oblig = NIL, External = NIL ) )

This lexical entry's $\theta$-grid has one syntactically obligatory argument, an agent NP. The other two arguments, a goal PP and a theme PP are optional, as is evidenced by the sentences in (159):

(159) a. John spoke.
b. John spoke to Mary.
c. John spoke about the problem.
d. John spoke to Mary about the problem.

Since the agent NP $\theta$-structure appears first in the $\theta$-grid, it marks the external or subject role to be assigned. The other two arguments – the goal and theme PP arguments – are the internal arguments in this $\theta$-grid.

### 4.4. The Parsing Algorithm

The left-corner parsing algorithm that I assume consists of two basic steps: 1) building tree structure(s) (nodes) for a lexical item based on its lexical entry and the grammar; and 2) attaching tree structures together. Section 4.4.1 describes what structures are built and predicted given a lexical entry. Section 4.4.2 describes how these structures are attached. Section 4.4.3 describes how gaps are posited in order to handle long distance dependencies. Section 4.4.4 gives the control structure of the underlying parser algorithm. Section 4.4.5 gives example parses using this parsing algorithm. Finally, Section 4.4.6 describes my assumptions regarding how reanalysis takes place if the parse fails because of the memory limitation constraints imposed upon it.

### 4.4.1. Node Projection

The parsing model assumed here performs morphological analysis on an input word to determine what root strings to look up in the lexicon. The result of lexical look-up and morphological analysis is a set of feature structures, each of which consists of a syntactic category with agreement features and a (possibly empty) $\theta$-grid. The morphological analysis tool that is used in the implemented parsing model is called DiMorph (Gibson (1991)). DiMorph is a morphological analyzer which accesses context-sensitive spelling change rules for English. However, it is not important with respect to this parsing model how morphological analysis takes place: all that is important is that such analysis (which includes lexical look-up) returns a set of feature structures representing all lexical entries for a word. See Swinney (1979), Tanenhaus, Leiman & Seidenberg (1979) and Seidenberg, Waters, Sanders & Langer (1984) among others for evidence of the psychological plausibility of such a process. Thus when the word *rock*, which has both nominal and verbal lexical entries, is input, both nominal and verbal entries for that word are accessed. The structures that are built from a given lexical entry consist of three components:

- The $\overline{\text{X}}$ structure of the category of the given lexical item.
- The $\overline{\text{X}}$ structure of categories whose heads are predicted to the right of the input word as dictated by left-corner prediction, where the prediction is guided by the $\theta$-grid of the given lexical item and the grammar of the language. These predicted categories mark the positions of the structures representing future input.
- The $\overline{\text{X}}$ structure of categories whose heads are to the left of the input word as indicated by the $\theta$-grid of the given lexical item.

Categories that are predicted to the right are expanded further in two cases: 1) if the predicted category matches the left-hand side of a rule that obligatorily applies in the grammar; and 2) if the predicted category matches the left-hand side of a rule whose left-corner category may be non-lexical. If one of these conditions is met, then the RHS categories of the rule in question are also hypothesized.

For example, an IP always has a subject noun phrase or Comp phrase, so a subject phrase is hypothesized at the specifier level having the disjunctive category consisting of noun or Comp.[68] Furthermore, the categories Comp and Infl can be non-lexical, so their complements, IP and VP respectively, are always hypothesized below the one-bar $\overline{\text{X}}$ level. Similarly, since relative clause operators may be non-lexical, they are always predicted below a relative clause CP.

In order to block the hypothesis of an infinite number of categories in a recursive set of rules, it is assumed that the rules to be considered in the expansion of RHS categories on

[68]The fact that all IPs have subjects is referred to as the Extended Projection Principle in Government-Binding theory (Chomsky (1981, 1982), *cf.* Rothstein (1983)).

a particular rule do not include that rule. Thus a rule may not be recursively predicted to the right at the structure hypothesis stage.

The structures predicted to the right of the given word will be referred to as *hypothesized* nodes (H-nodes). All other structures will be called *confirmed* nodes (C-nodes). H-nodes are further subdivided into two classes: optional H-nodes and obligatory H-nodes. Obligatory H-nodes are H-nodes whose presence with respect to some C-node is forced, as determined by the lexicon or $\overline{\mathrm{X}}$ rule. For example, adjunct $\overline{\mathrm{X}}$ rules never obligatorily apply. Thus H-nodes proposed from such rules are marked optional. Lexical entries, on the other hand, can obligatorily require certain arguments. For example, one lexical entry for the verb *give* requires the presence of three arguments. Thus the H-nodes that are proposed corresponding to these arguments are obligatory ones. Optional H-nodes are simply pruned from a structure at the end of a parse.

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. Section 4.4.2 gives a description of how attachments take place.

It is assumed for implementation efficiency reasons that all of the structures that are built from the lexical entries for a word are pre-compiled in the lexicon. Hence obtaining $\overline{\mathrm{X}}$ structures from the lexicon involves only table look-up, without any additional computation. However, no empirical effects rely on this assumption; this assumption is made only to allow more efficient parsing. Furthermore, in order to avoid a possible space shortage, it is assumed that structures are shared to a great degree in the lexicon. For example, two lexical entries with the same syntactic category contain pointers to only one set of relevant structures; these structures are not repeated for each of the two entries (*cf.* Flickinger, Pollard & Wasow (1985)).

Examples will help make clear what structures are accessed when a word is input. First consider the determiner *the*. By $\overline{\mathrm{X}}$ Theory, a two-bar projection for this category is constructed. In order to see what further structures are built around this base structure we must examine this lexical entry's $\theta$-grid as well as the grammar rules that involve its category. The lexical entry for the determiner *the* contains an empty $\theta$-grid, and the only rule in the grammar which contains a determiner is as follows:

(160) NP $\rightarrow$ DetP N$'$

Since a determiner phrase appears in the left corner of the above NP rule, an H-node NP is constructed above the DetP. The head of this NP lies to the right of the C-node DetP. Since there are no NP rules in the grammar that obligatorily apply or whose left-corner is optionally non-lexical, no further structures can be hypothesized. The structure in (161) is thus the only structure that is built when the determiner *the* is input:[69]

[69]An empty string which is immediately dominated by a zero-level H-node will be represented as $h$ to indicate its hypothesized status. An empty string which is immediately dominated by a zero-level C-node will be represented as $e$.

(161) $[_{NP}\ [_{DetP}\ [_{Det'}\ [_{Det}$ the $]]]\ [_{N'}\ [_N\ h\ ]]]$

Now consider the noun *rock*. As with all categories, a two bar projection is built for this category. The nominal lexical entry for *rock* has no $\theta$-grid, but many rules in the grammar involve nouns. Some of these rules are given below:

(162) a. IP $\rightarrow$ {NP, CP} I′
b. NP $\rightarrow$ DetP N′
c. $N'_1 \rightarrow N'_2$ PP
d. $N'_1 \rightarrow N'_2$ CP

Rule (162a) states that an NP may act as specifier for an IP. Since a projection of a noun is in the left corner of this rule, the LHS category is projected above the base NP structure, and an $\overline{\mathrm{X}}$ structure is built for the category Infl as well. This category is further expanded to the right, since the head (the left corner) of this category can be non-lexical. Hence we have the following structure thus far:

(163) $[_{IP}\ [_{NP}\ [_{N'}\ [_N$ rock $]]]\ [_{I'}\ [_I\ h\ ]\ [_{VP}\ [_{V'}\ [_V\ h\ ]]]\ ]]$

Consider now rules (162b)–(162d) with respect to the hypothesis of categories from a C-node noun phrase. Rule (162b) dictates that a noun phrase can consist of a determiner phrase followed by the category N′. Although this rule deals with the expansion of noun phrases, it does not contain a projection of a noun in its left-corner, so it does not apply here. Only a determiner will trigger this rule.

Rules (162c) and (162d) both contain a nominal projection in their left corners, and so each of these rules causes prediction of categories to the right here. First consider (162c) in which the category N′ can be expanded as a further N′ followed by a prepositional phrase. As a result of this rule, a hypothesized PP node is adjoined to the N′ level. Rule (162d) causes a similar adjunction and attachment of CP relative clause. Since the specifier of CP relative clause can be non-lexical, this category is also predicted. Furthermore the complement of the CP is predicted, an IP, and this IP is also expanded. Thus CP and PP modifying phrases are hypothesized at the N′ level, modifying the NP headed by *rock*. In order to reduce the space involved in H-node projection, I assume that these two hypothesized nodes are attached in the same structure. This joint attachment will be notated by prefixing the special symbol *or* before the hypothesized structures:[70]

(164) $[_{NP}\ [_{N'}\ [_{N'}\ [_N$ rock $]]$ (*or* $[_{PP}\ h\ ]\ [_{CP}\ [_{NP}\ h\ ]\ [_C\ h\ ]\ [_{IP}\ [_{NP}\ h\ ]\ [_I\ h\ ]\ [_{VP}\ h\ ]]]$ ) $]]$

Other categories that can be hypothesized at the N′ level are assumed to be attached in the same way.

Hence the noun *rock* is associated with structures consisting of the combination of the structures in (163) and (164). That is, the structure depicted in (163) is built above that in

[70]From now on I will only depict the maximal projections of H-nodes, skipping over inner bar-levels. However, keep in mind that the inner bar-levels are still assumed to exist.

(164), to give (165):

(165) $[_{IP}\ [_{NP}\ [_{N'}\ [_{N'}\ [_{N}\ \text{rock}\ ]]\ (\text{*or*}\ [_{PP}\ h]\ [_{CP}\ [_{NP}\ h\ ]\ [_{C}\ h\ ]\ [_{IP}\ [_{NP}\ h\ ]\ [_{I}\ h\ ]\ [_{VP}\ h\ ]]]\ )\ ]]$
$[_{I}\ h\ ]\ [_{VP}\ h\ ]\ ]]]$

The word *rock* has verbal lexical entries in addition to its nominal lexical entries. There are at least two possible subcategorization patterns for this verb: one requiring an external NP agent and an internal NP theme, and a second requiring only an external NP theme.[71] Since the external thematic role is assigned to the left in English, a lexically empty NP is built and attached as specifier of the VP, to the left of the head verb (see Fukui & Speas (1986), Pollard & Sag (1987) and the references cited in each). Since this structure is not to the right of all lexical material, it is not an H-node and hence does not mark a possible attachment site (see Section 4.4.2.) Since internal subcategorizations appear to the right in English, H-node structures for internal subcategorization patterns are built and placed below the V′ projection for the reading of *rock* that has internal subcategorization requirements. Furthermore, the rules in the grammar allow modification of verbs at the V′ level by PPs and CPs (possibly among other categories).

The two subcategorization patterns of the infinitival form of *rock* are represented in (166). The structures below the hypothesized modifying CP node are left out to make the structure as a whole easier to view.

(166) a. $[_{VP}\ [_{NP}\ e]\ [_{V'}\ [_{V'}\ [_{V}\ \text{rock}\ ]]\ (\text{*or*}\ [_{PP}\ h\ ]\ [_{CP}\ h\ ])\ ]]$
b. $[_{VP}\ [_{NP}\ e]\ [_{V'}\ [_{V'}\ [_{V}\ \text{rock}\ ]\ [_{NP}\ h\ ]]\ (\text{*or*}\ [_{PP}\ h\ ]\ [_{CP}\ h\ ])\ ]]$

The word *rock* can also indicate a tensed verb with non-third person singular agreement features. IP representations for each subcategorization of *rock* are therefore constructed, giving the following additional structures:

(167) a. $[_{IP}\ [_{I'}\ [_{I}\ +tense\ ]\ [_{VP}\ [_{NP}\ e\ ]\ [_{V'}\ [_{V'}\ [_{V}\ \text{rock}\ ]]\ (\text{*or*}\ [_{PP}\ h\ ]\ [_{CP}\ h\ ])\ ]]]]$
b. $[_{IP}\ [_{I'}\ [_{I}\ +tense\ ]\ [_{VP}\ [_{NP}\ e\ ]\ [_{V'}\ [_{V'}\ [_{V}\ \text{rock}\ ]\ [_{NP}\ h]]\ (\text{*or*}\ [_{PP}\ h\ ]\ [_{CP}\ h\ ])\ ]]]]$

Thus all the structures in (168) are accessed when the word *rock* is encountered:

(168) a. $[_{IP}\ [_{NP}\ [_{N'}\ [_{N'}\ [_{N}\ \text{rock}\ ]]\ (\text{*or*}\ [_{PP}\ h]\ [_{CP}\ [_{NP}\ h\ ]\ [_{C}\ h\ ]\ [_{IP}\ [_{NP}\ h\ ]\ [_{I}\ h\ ]\ [_{VP}\ h\ ]]]\ )\ ]]$
$[_{I}\ h\ ]\ [_{VP}\ h\ ]\ ]]]$
b. $[_{VP}\ [_{NP}\ e]\ [_{V'}\ [_{V'}\ [_{V}\ \text{rock}\ ]]\ (\text{*or*}\ [_{PP}\ h\ ]\ [_{CP}\ h\ ])\ ]]$
c. $[_{VP}\ [_{NP}\ e]\ [_{V'}\ [_{V'}\ [_{V}\ \text{rock}\ ]\ [_{NP}\ h\ ]]\ (\text{*or*}\ [_{PP}\ h\ ]\ [_{CP}\ h\ ])\ ]]$
d. $[_{IP}\ [_{I'}\ [_{I}\ +tense\ ]\ [_{VP}\ [_{NP}\ e\ ]\ [_{V'}\ [_{V'}\ [_{V}\ \text{rock}\ ]]\ (\text{*or*}\ [_{PP}\ h\ ]\ [_{CP}\ h\ ])\ ]]]]$
e. $[_{IP}\ [_{I'}\ [_{I}\ +tense\ ]\ [_{VP}\ [_{NP}\ e\ ]\ [_{V'}\ [_{V'}\ [_{V}\ \text{rock}\ ]\ [_{NP}\ h]]\ (\text{*or*}\ [_{PP}\ h\ ]\ [_{CP}\ h\ ])\ ]]]]$

---

[71] One of these lexical entries is probably lexically derivable from the other via a lexical rule. However, the implementation of the lexicon that I assume includes no such lexical rules yet, although it would not be difficult in principle to add such rules.

Consider now the verb *know*, which is assumed to have at least three lexical entries, all of which subcategorize for an external NP experiencer, but which vary with respect to the internal subcategorized elements: one subcategorizes for an internal theme NP, the second subcategorizes for an internal proposition CP and the third for an internal proposition IP. Thus the structures built for the infinitival form of the verb *know* are as in (169):

(169) a. $[_{VP}$ $[_{NP}$ $e$ ] $[_{V'}$ $[_{V'}$ $[_{V}$ know ] $[_{NP}$ $h$ ]] (*or* $[_{PP}$ $h$ ] $[_{CP}$ $h$ ]) ]]
b. $[_{VP}$ $[_{NP}$ $e$ ] $[_{V'}$ $[_{V'}$ $[_{V}$ know ] $[_{CP}$ $h$ ]] (*or* $[_{PP}$ $h$ ] $[_{CP}$ $h$ ]) ]]
c. $[_{VP}$ $[_{NP}$ $e$ ] $[_{V'}$ $[_{V'}$ $[_{V}$ know ] $[_{IP}$ $h$ ]] (*or* $[_{PP}$ $h$ ] $[_{CP}$ $h$ ]) ]]

Consider now the adjective *big*. This adjective has no $\theta$-grid, but there exists a rule in the grammar with an adjectival phrase in its left corner:

(170) $N'_1 \rightarrow AP\ N'_2$

As a result of this rule, an H-node N′ node is constructed above every adjective, so that the adjective *big* is associated with the following structure:

(171) $[_{NP}$ $[_{N'}$ $[_{AP}$ $[_{A'}$ $[_{A}$ big ]]] $[_{N'}$ $[_{N}$ $h$ ]]]]

### 4.4.2. Node Attachment

As a result of the local predictions, attachment in the framework proposed here consists of matching H-nodes against the nodes that occur next in the input stream. This matching is simply a *unification* of the features of the hypothesized nodes against those of nodes to follow.[72] For example, suppose that the word *the* is on a stack as a determiner phrase which makes up the specifier of an H-node noun phrase. Furthermore, suppose that the word *big* is represented in the buffer as an adjective phrase attached as a pre-head modifier of an H-node noun phrase:

(172)

Stack: ( { $[_{NP}$ $[_{DetP}$ $[_{Det'}$ $[_{Det}$ the ]]] $[_{N'}$ $[_{N}$ $h$ ]]] } )
Buffer: { $[_{NP}$ $[_{N'}$ $[_{AP}$ $[_{A'}$ $[_{A}$ big ]]] $[_{N'}$ $[_{N}$ $h$ ]]]] }

At this point the attachment algorithm will allow the H-node NP on the stack to unify with the H-node NP in the buffer to give the structure in (173):

(173) $[_{NP}$ $[_{DetP}$ the ] $[_{N'}$ $[_{AP}$ big ] $[_{N'}$ $[_{N}$ $h$ ]]]]

Suppose the next word in the input stream is the word *rock*, which has both nominal and verbal readings as depicted in the previous section in (168). While the verbal readings of

[72]See Sheiber (1986) for a formal definition of unification as well as background on possible uses for unification in particular grammar formalisms.

*rock* are not compatible with the NP H-node in (173), the nominal reading of *rock* can be attached as the head noun in (173). The resulting structure is given in (174):

(174) $[_{IP}$ $[_{NP}$ $[_{DetP}$ the ] $[_{N'}$ $[_{AP}$ big ] $[_{N'}$ $[_{N'}$ $[_N$ rock ]] (*or* $[_{PP}$ $h]$ $[_{CP}$ $h$ ]) ]] $[_I$ $h$ ]]

Attachment is formally defined as follows (*cf.* Gibson (1989)):

A structure $B$ in the buffer can attach to a structure $A$ on top of a stack iff all of the following are true:

1. Structure $A$ contains an H-node structure, *stack-site*, which dominates no lexical material.
2. Structure $B$ contains a node, *buffer-site*, that dominates all zero-level C-nodes (and hence all lexical material) in $B$.
3. The lexical features and bar-levels of structures *stack-site* and *buffer-site* are compatible.
4. The tree structure above *stack-site* is compatible with the tree structure above *buffer-site*, up to the existence of optional H-nodes.

If attachment is viable, copies of the parent tree structures of *stack-site* and *buffer-site* are unified to give the resulting structure.

Since the order of the words in the input must be maintained in a final parse, only those nodes in a buffer structure that dominate all lexical items in that structure are permissible as attachment sites. For example, suppose that the buffer contained a representation for the noun phrase *dogs in parks*. Furthermore, suppose that there is an H-node NP on a stack representing the word *the*. Although it would be suitable for the buffer structure representing the entire noun phrase *dogs in parks* to match the stack H-node, it would not be suitable for the C-node NP representing *parks* to match this H-node. This attachment would result in a structure that moved the lexical input *the* to the right of the words *dogs in*, producing a parse for the input *dogs in the park*. Since the word order of the input string must be maintained, sites for buffer attachment must dominate all lexical items in the buffer structure.

Parent tree structure unification is made up of the following processes:

- Copy *stack-site* and *buffer-site* and their respective parent tree structures. Label the tree structure corresponding to *stack-site* as *new-stack-site*, and the tree structure corresponding to *buffer-site* as *new-buffer-site*.
- Set the lexical features of *stack-site* and *buffer-site* to be the unification of their lexical features.
- Attach the children of *buffer-site* in the position previously occupied by the rightmost child of *stack-site*.

- Set the tree structure above the maximal projection of *stack-site* to be the unification of the parent tree structures of the maximal projection of *stack-site* and the maximal projection of *buffer-site*.

For example, consider the potential attachment of structure (175) in the buffer to structure (176) on top of a stack:[73]

(175) $[_{IP}$ $[_{NP}$ $[_{N'}$ $[_{N}$ John ]]] $h$ ]

(176) $[_{IP}$ $[_{NP}$ Mary$_i$ ] $[_{VP}$ $[_{NP}$ $e_i$] $[_{V'}$ $[_{V'}$ $[_{V}$ sees ] $[_{NP}$ $h$ ]] (*or* $[_{PP}$ $h$] $[_{CP}$ $h$ ]) ]]]

The H-node NP in object position of the structure in (176) (the stack-site) is compatible with the C-node NP which makes up part of the structure in (175) (the buffer-site). Since the category and bar-level of these two structures are compatible, and furthermore since the categories dominating the C-node NP *John* in (175) are optional (that is, they are not obligatory with respect to the existence of the NP *John*), the attachment is viable. The resulting unification gives the structure in (177):

(177) $[_{IP}$ $[_{NP}$ Mary$_i$ ] $[_{VP}$ $[_{NP}$ $e_i$ ] $[_{V'}$ $[_{V'}$ $[_{V}$ sees ] $[_{NP}$ John ]] (*or* $[_{PP}$ $h$] $[_{CP}$ $h$ ]) ]]]

While the NP H-node in (176) and the NP C-node in (175) categories in the above attachment can serve as appropriate stack and buffer sites for attachment, in fact any of the $\overline{\text{X}}$ levels inside these NPs – the NP, N′ or N level – would serve as appropriate sites. For example, the H-node of category N in (176) along with the C-node of category N in (175) would also serve as appropriate attachment sites according to the attachment algorithm given above. Furthermore, this attachment would result in a structure identical to that in (177). As a result of the multiple derivation possibility, I hypothesize that attachments between projections of two head categories only be tried until there is a success. Once there is successful attachment, there is no need to try further attachments between the two projections. In practice, the implementation of attachment attempts to match two-bar projections first, then one-bar projections, then zero-level projections. If a match between two nodes succeeds at one level, then additional attempts between the same $\overline{\text{X}}$ projections are not be attempted at lower levels.

Consider now a further example, this one involving adjunct attachment. Given the structure in (177) on top of a stack, suppose that the preposition *in* is input into the buffer, as depicted in (178):

(178) $[_{PP}$ $[_{P'}$ $[_{P}$ in ] $[_{NP}$ $h$ ]]]

This structure matches the modifier PP H-node in (177) and hence attachment can take place. However, modifier attachment is slightly different from other kinds of attachment since any number of the hypothesized categories can modify a head category. That is, even if an H-node is matched and a C-node is attached in its place, it is still grammatically

[73]See Section 4.4.3 for a description of how the IP subject *Mary* gets coindexed with the VP subject trace $e_i$ in (176).

possible to attach further nodes of the same category, adjoined at that $\overline{X}$ level. For example, it grammatical to attach two PPs as a modifier of a VP, as in (179):

(179) Mary saw John in the park on Tuesday.

The structure for such a sentence under the binary branching $\overline{X}$ modifier rules assumed here is as in (180):

(180) [$_{IP}$ [$_{NP}$ Mary$_i$ ] [$_{VP}$ [$_{NP}$ $e_i$ ] [$_{V'}$ [$_{V'}$ [$_{V'}$ [$_V$ saw ] [$_{NP}$ John ]] [$_{PP}$ in the park ]] [$_{PP}$ on Tuesday ]]]]

In order to allow multiple adjuncts at the same $\overline{X}$ level, it is necessary to adjoin the prospective adjuncts to their attachment sites. That is, if a structure matches an H-node modifying phrase, then that structure is adjoined above the attachment site so that future structures can also attach as a modifier to the same head. Thus the attachment of (178) which represents *in* to (177) which represents *Mary saw John* results in the following structure:

(181) [$_{IP}$ [$_{NP}$ Mary$_i$ ] [$_{VP}$ [$_{NP}$ $e_i$ ] [$_{V'}$ [$_{V'}$ [$_{V'}$ [$_V$ sees ] [$_{NP}$ John ]] [$_{PP}$ [$_{P'}$ [$_P$ in ] [$_{NP}$ $h$ ]]]] (*or* [$_{PP}$ $h$] [$_{CP}$ $h$ ]) ]]]

In particular, the PP *in* adjoins to the V′ level of the VP, resulting in an additional adjunction level. Further modifiers are still proposed at the higher V′ level.

### 4.4.3. Chain Formation

I assume the existence of the data structure *chain* (Chomsky (1981)) which links phrases in positions that are not interpretable to positions that allow interpretation: either positions that receive thematic roles or adjunct positions. A chain is defined here in terms of two types of *sub-chains*: *simple sub-chains* and *complex sub-chains*. A *simple sub-chain* consists of a pointer to a structure whose position can be interpreted together with the structural head that provides the interpretation for that position: either as an argument in which a thematic role is assigned or as an adjunct. The structure component of each sub-chain will be called the *head* of that sub-chain. A *complex sub-chain* is recursively defined to consist of a pointer to a structure with agreement features - the complex sub-chain's head - along with a further sub-chain - the complex sub-chain's *tail*. Furthermore, it is assumed: 1) that the category and agreement features of the head of a complex sub-chain agree in category and agreement features with the head of the tail of the complex sub-chain; and 2) that the head of a complex sub-chain is subjacent to the head of the tail of the complex sub-chain (Chomsky (1973, 1981, 1986b)). The head of a complex sub-chain is said to be *coindexed* with the head of its tail.

A *chain* consists of a sub-chain whose head has thematic content and whose tail is a *trace sub-chain*. A trace sub-chain (or gap sub-chain) is defined to be a sub-chain whose head is non-lexical and whose tail either is a further trace sub-chain or is null (in which case, the trace sub-chain is a simple sub-chain).

Furthermore, following Chomsky (1981), I assume the existence of a further division of sub-chains: *A-sub-chains* and *$\overline{A}$-sub-chains*. An A-sub-chain is a sub-chain whose head is in an argument position (A-position: sisters to the lexical head and IP specifiers) and whose tail (if it exists) is a further A-sub-chain. An $\overline{\text{A}}$-sub-chain consists of a sub-chain whose head is in the non-argument position spec(CP) and whose tail is a sub-chain (either an A-sub-chain or a further $\overline{\text{A}}$-sub-chain).

Finally, an *A-chain* is defined to be a chain which is also an A-sub-chain, while an *$\overline{A}$-chain* is defined to be a chain which is also an $\overline{\text{A}}$-sub-chain.

Examples of chains will help clarify these definitions. Consider the structure in (182):

(182) $[_{IP}$ $[_{NP}$ Mary$_i$ ] $[_{VP}$ $[_{NP}$ $e_i$ ] $[_{V'}$ $[_{V}$ saw ] $[_{NP}$ John ]]]]

The NP headed by *John* forms the head of a simple A-chain which receives its interpretation directly from the verb *saw*. Furthermore, the NP *Mary* makes up the head of a complex A-chain whose tail is a simple sub-chain that receives its thematic role from the verb *saw*.

Consider now (183):

(183) $[_{CP}$ $[_{NP}$ who$_i$ ] $[_{C}$ did ] $[_{IP}$ $[_{NP}$ Mary$_j$ ] $[_{VP}$ $[_{NP}$ $e_j$ ] $[_{V'}$ $[_{V}$ see ] $[_{NP}$ $e_i$ ] ]]]]]

As in (182), the NP *Mary* heads a complex A-chain whose tail is a simple A-sub-chain that receives its thematic role from the verb *see*. Furthermore, the NP *who* heads a complex $\overline{\text{A}}$-chain whose tail is a simple A-sub-chain that receives its thematic role from the verb *see*.

Finally consider (184):

(184) $[_{CP}$ $[_{NP}$ who$_i$ ] $[_{IP}$ $[_{NP}$ $e_i$ ] $[_{VP}$ $[_{NP}$ $e_i$ ] $[_{V'}$ $[_{V}$ saw ] $[_{NP}$ John ]]]]

As in (182), the NP *John* heads a simple A-chain which receives a thematic role from the verb *saw*. The NP *who* heads a complex $\overline{\text{A}}$-chain whose tail is a complex A-sub-chain. The tail of this sub-chain is a simple A-sub-chain whose head receives its thematic role from the verb *saw*.

Once attachments have taken place at a parse state, the process of *chain extension* is applied to the resultant structures. $\overline{\text{A}}$-chain extension is a local operation that applies to a structure if that structure includes a constituent in the specifier of CP (*cf.* Gibson & Clark (1987)). Given the existence of such a constituent, trace categories are then placed in further CP specifier positions, or are placed in available argument positions – those currently occupied by lexically empty H-nodes – that are compatible in agreement features (person, number, case *etc.*) with the chain's head. Moreover, this gap-positing process is assumed to adhere to the principle of Subjacency (Chomsky (1973, 1981), *cf.* Chomsky (1986b)). This principle restricts the placement of successive traces so that they are not separated by more than one bounding node, where the bounding nodes are taken to be CP and NP in English.

Similarly, A-chain extension takes place given a constituent in the specifier position of IP that does not receive a thematic role. A-chain traces are placed in further A-positions until a thematic role is located for the chain. A-chain extension is also constrained by Subjacency (see Chomsky (1986b) and the references cited there).

For example, consider the structure in (185), the result of attachment of a structure representing the word *who* to a structure representing the word *saw*:

(185) $[_{CP}$ $[_{NP}$ who ] $[_{IP}$ $[_{NP}$ $e$ ] $[_{VP}$ $[_{NP}$ $e$ ] $[_{V'}$ $[_{V}$ saw ] $[_{NP}$ $h$ ] ]]]]

The *wh*-NP *who* requires a thematic role, but does not receive one in this structure, so $\overline{\text{A}}$-chain extension takes place. As a result, the trace in the subject position of the matrix IP is coindexed with this NP:

(186) $[_{CP}$ $[_{NP}$ who$_i$ ] $[_{IP}$ $[_{NP}$ $e_i$ ] $[_{VP}$ $[_{NP}$ $e$ ] $[_{V'}$ $[_{V}$ saw ] $[_{NP}$ $h$ ] ]]]]

The gap in IP subject position does not receive a thematic role so A-chain extension takes place, resulting in an additional chain linking the specifier of IP to the specifier of VP:

(187) $[_{CP}$ $[_{NP}$ who$_i$ ] $[_{IP}$ $[_{NP}$ $e_i$ ] $[_{VP}$ $[_{NP}$ $e_i$ ] $[_{V'}$ $[_{V}$ saw ] $[_{NP}$ $h$ ] ]]]]

Since the verb *saw* assigns a thematic role to this position, no further chain extension takes place, and the chain is complete.

Consider now the structure in (188), representing the (partial) question *who did Mary believe*:

(188) $[_{CP}$ $[_{NP}$ who ] $[_{C}$ did ] $[_{IP}$ $[_{NP}$ Mary$_j$ ] $[_{VP}$ $[_{NP}$ $e_j$ ] $[_{V'}$ (*or* ($[_{V}$ believe ] $[_{NP}$ $h$ ]) ($[_{V}$ believe ] $[_{CP}$ $h$ ])) ]]]]

The *wh*-NP *who* requires a thematic role, but doesn't receive one in this structure, so the process of chain extension is initiated. The verb *believe* is ambiguous, so that two possible mutually exclusive chain extensions are possible: one in the object position of one reading of *believe* and one in the specifier position of the CP complement of another reading of *believe*. These two chain extensions are given as follows:

(189) $[_{CP}$ $[_{NP}$ who$_i$ ] $[_{C}$ did ] $[_{IP}$ $[_{NP}$ Mary$_j$ ] $[_{VP}$ $[_{NP}$ $e_j$ ] $[_{V'}$ (*or* ($[_{V}$ believe ] $[_{NP}$ $e_i$ ]) ($[_{V}$ believe ] $[_{CP}$ $[_{NP}$ $e_i$] $[_{C'}$ $h$ ])) ]]]]

The gap in the object position of *believe* completes one chain, and thus no further extension is attempted on this chain. However, the gap in the specifier position of the CP complement to *believe* does not complete its chain, and further extension is necessary. For example, this chain can be continued as in (190):

(190) Who did Mary believe that John saw?

### 4.4.4. The Control Structure of the Underlying Parser

The underlying parsing algorithm that I propose here is quite simple: attachments are attempted between structures in the stack-set and structures in the buffer, while, in parallel, a copy of the contents of the buffer is pushed onto a copy of each stack of the stack-set. This algorithm is constrained by the memory limitation constraints given in Chapter 1 (and developed in Chapters 5 and 7) as well as the Stack Push Constraint, which limits the number of possible stack push operations. The Stack Push Constraint is motivated by my assumption regarding human linguistic processing breakdown – the Single Cell Hypothesis:

(191) The Single Cell Hypothesis (SCH):
Human linguistic processing breakdown occurs if there are no single cell representations for the current input after attachments have been made.

Thus if a word is input into the buffer and no attachments are possible with any of the stacks in the stack-set, then processing breakdown has occurred.[74,75] The Single Cell Hypothesis motivates the Stack Push Constraint, which prefers attached structures over unattached structures. The Stack Push Constraint (SPC) is given in (192):

(192) The Stack Push Constraint:
Push onto stacks only those structures whose root nodes are H-nodes and which can't attach to any structure.

Thus if a structure in the buffer can attach to a structure on a stack, then that attachment is made, and the buffer structure is not pushed onto any stacks. Furthermore, of those buffer structures that cannot attach to any stack structure, only H-nodes are pushed onto stacks, since it turns out that these are the only structures that can possibly end up as part of single cell representations for the input string. The proof of this assertion follows its restatement in (193):

(193) Stack Push Constraint Theorem:
If a node in the buffer cannot attach to any stack structure, then its root node must be an H-node if it is ever going to be part of a single cell representation for the input string.

> *Proof*: Suppose there exists a buffer node $A$ that cannot attach to any structure on top of any stack in the current stack-set. Furthermore suppose that that the root node of $A$ is a C-node. It will be proved that $A$ cannot end up as part of a single cell representation for the input string.
>
> *Case 1*: Since $A$ cannot attach to any node on top of a current stack, $A$ is pushed onto to each of these stacks (along with other nodes that can't attach), to form

[74] I also assume that reanalysis is attempted once processing breakdown has occurred. My assumptions regarding reanalysis are given in Section 4.4.6.

[75] While the SCH hypothesizes that breakdown occurs if there are no single cell representations for the input, it does not mean that only single cell stacks are pursued at each parse state. Multiple cell stacks can continue to exist as long as there is at least one single cell stack in existence as well.

a set of new stacks. Attachments are then attempted between a new buffer representing the next word in the input and these stacks. Suppose $A$ contains no H-nodes at all. Consider any one of these stacks on top of which is $A$. Since an H-node is required for attachment to a stack node, $A$ cannot take part in attachment, since $A$ contains no H-nodes. Furthermore $A$ will never be able to attach to future structures since it contains no H-nodes. Since $A$ cannot attach to material below it on the stack by hypothesis and it cannot attach to material following it, $A$ will never attach to another structure, and the desired conclusion is satisfied.

*Case 2*: Suppose $A$ contains an H-node $H$. Consider any one of these stacks on top of which is $A$, along with an arbitrary node $B$ in the buffer.

*Case 2.1*: Suppose that $A$ and $B$ cannot attach. Then we reach the desired conclusion as above.

*Case 2.2*: Suppose that $A$ and $B$ can attach. By the definition of attachment, the buffer attachment site $B'$ (part of $B$) that matches $H$ must dominate all zero-level C-nodes in $B$. Thus the only zero-level nodes in $B$ that the matching node $H$ might not dominate inside $B$ are H-nodes.

But H-nodes all occur to the right of all C-nodes in any structure, by definition. Thus these H-nodes will not be unifiable with C-nodes in the stack node $A$, since all C-nodes in the stack node $A$ occur to the left of $H$, the site of attachment in $A$.

Hence all structure attachment takes place below all C-nodes in $A$, and therefore no structures are attached above the root node of the stack node $A$, a C-node. Thus the features of the $A$ are not altered by attachment.

Since $A$ can't attach to structures below it on the stack (by hypothesis) and it can never change, it will therefore never be able to attach. Thus $A$ will never be part of a single cell representation for the input string. *Q.E.D.*

Since buffer structures that cannot currently attach to a stack structure and whose root nodes are C-nodes can never be part of a single cell representation for the input string, these structures are not retained; the Stack Push Constraint does not push them onto stacks.

The Stack Push Constraint is related to the Exclusive Attachment Constraint (EAC) of Gibson (1987), Gibson & Clark (1987), Clark & Gibson (1988). The EAC states that if attachments are possible at any parse state, then all structures that cannot take part in attachment are pruned. Note that the EAC is therefore quite different from the SPC, since the SPC *saves* structures that cannot immediately take part in attachment, but looks to avoid maintaining redundant structures, while the EAC *removes* structures that cannot immediately take part in attachments.[76]

[76]The EAC was abandoned because of its empirical inadequacy. For example, the EAC predicted garden-path effects on locally ambiguous sentences such as those in (194), an undesirable consequence.

(194)a. Bill knew John liked Mary.

b. John saw the light beer can fall over.

The underlying parsing algorithm proposed here (including the Stack Push Constraint) is given below in pseudo-code, as well as in Figure 10 in flow chart form.

1. (Initializations) Set BUFFER to the empty set. Set STACK-SET to be the set containing an empty stack. Set REMAINING to be the input string.

2. If REMAINING is empty, then return STACK-SET. Stop.
   Else do the following:

   (a) Set CURRENT-WORD to be the first word in REMAINING.

   (b) Set REMAINING to be the rest of the words in REMAINING.

   (c) Build all maximal projection nodes corresponding to CURRENT-WORD and place them in BUFFER. (See Section 4.4.1 on the projection of nodes from the lexicon.)

3. Set the updated STACK-SET to be all the stacks returned by calling the function **Perform-All-Operations** on each STACK and BUFFER combination, where STACK is a member of STACK-SET.

4. Go to 2.

Figure 10: Flowchart for the Underlying Parsing Algorithm

The definition of the recursive function **Perform-All-Operations** is given below:

**Perform-All-Operations** (BUFFER, STACK)

1. Set ATTACHMENTS to be all possible attachments between the buffer and the top of the stack.

2. Return the set of stacks consisting of:

   - The stack formed by pushing the H-nodes in BUFFER that do not take part in attachments on top of STACK. (This step has the Stack Push Constraint built into it.)
   - (If ATTACHMENTS is non-null) The stacks returned by calling the function **Perform-All-Operations** on ATTACHMENTS and the stack that results from popping STACK.

The constraints on memory capacity which are outlined in Chapters 1, 5 and 7 apply to the representations that this algorithm produces in order to give what is hypothesized to be a psychologically plausible parser.

### 4.4.5. Example Parses

This section contains two example parses using the parsing algorithm given in the previous section. The first example sentence to be parsed here is the simple sentence in (195):

(195) John likes Mary.

First the word *John* is read and an H-node IP for this word is placed in the buffer:

(196)

Buffer:
{ $[_{IP}$ $[_{NP}$ John ] $[_{I'}$ $[_{I}$ $h$ ] $[_{VP}$ $h$ ] ]] }

No attachments are possible between the buffer and the stack-set, since the only stack in the stack-set is empty. Thus the only possible operation at this parse state is a stack push. This operation is performed, resulting in a stack-set containing a single stack which contains the single buffer cell in (196). The next word *likes* is now read. A C-node Infl phrase is built for this word since it is a verb with tense information:

(197)

Buffer:
{ $[_{IP}$ $[_{I'}$ $[_{I}$ $+tense$ ] $[_{VP}$ $[_{NP}$ $e$] $[_{V'}$ $[_{V'}$ $[_{V}$ likes ] $[_{NP}$ $h$ ]] (*or* $[_{PP}$ $h$ ] $[_{CP}$ $h$ ]) ]]]] }

Attachments are now attempted between the buffer and the top of the solitary stack in the stack-set. The IP in the buffer matches the H-node Infl in the stack structure, resulting in the attachment of the NP *John* as subject of the verb *likes*:

(198) $[_{IP}$ $[_{NP}$ John ] $[_{I'}$ $[_I$ *+tense* ] $[_{VP}$ $[_{NP}$ *e* ] $[_{V'}$ $[_{V'}$ $[_V$ likes ] $[_{NP}$ *h* ]] (*or* $[_{PP}$ *h* ] $[_{CP}$ *h* ]) ]]]]

Since the buffer structure in (197) was able to attach to stack structures, it does not get pushed onto a stack because of the Stack Push Constraint.

At this point gap-positing is initiated, resulting in a chain that links the NP *John* in (198) to the specifier of the VP headed by *likes*.

The word *Mary* is then input to the buffer, giving a structure exactly like that for *John*:

(199)

Buffer:
{ $[_{IP}$ $[_{NP}$ Mary ] $[_{I'}$ $[_I$ *h* ] $[_{VP}$ *h* ] ]] }

This structure attaches as object of the verb *likes* in the single representation on top of the stack. This attachment does not maintain the H-nodes in (199), so these structures are pushed onto the current stack. Thus there are two stacks in the resulting stack-set: one which contains single cell representations of the input, and one which contains two cell representations of the input:

(200)

Stack-set:
( { $[_{IP}$ $[_{NP}$ $\text{John}_i$ ] $[_{VP}$ $[_{V'}$ $[_{V'}$ $[_V$ likes ] $[_{NP}$ Mary ]] (*or* $[_{PP}$ *h* ] $[_{CP}$ *h* ]) ]]] } )
( { $[_{IP}$ $[_{NP}$ Mary ] $[_{I'}$ $[_I$ *h* ] $[_{VP}$ *h* ] ]] }
{ $[_{IP}$ $[_{NP}$ $\text{John}_i$ ] $[_{VP}$ $[_{V'}$ $[_{V'}$ $[_V$ likes ] $[_{NP}$ *e* ]] (*or* $[_{PP}$ *h* ] $[_{CP}$ *h* ]) ]]] } )

The parse stops at this point since there are no further words in the sentence. Only single cell representations are returned, and optional H-nodes along with redundant $\overline{\text{X}}$ levels are pruned resulting in the following structure to be returned:

(201) $[_{IP}$ $[_{NP}$ $\text{John}_i$ ] $[_{I'}$ $[_I$ *+tense* ] $[_{VP}$ $[_{NP}$ $e_i$ ] $[_{V'}$ $[_V$ likes ] $[_{NP}$ Mary ]]] ]]

The second example sentence to be parsed is the sentence in (202):

(202) John believes the light man's friend ate the squid.

The parse of this sentence demonstrates the ability of the proposed parsing algorithm to maintain multiple representations and multiple stacks in situations where that either of a number of possible readings can be continued without processing difficulty.

The parse of the words *John believes* is analogous to the parse of *John likes* in the previous example, except that the verb *believes* can take three different phrasal arguments: a theme

NP, a proposition CP or a proposition IP. Thus the stack-set consists of a single stack containing three IP readings for the input *John believes*, as depicted in (203):[77]

(203)

Stack-set:
$( \ \{ \ [_{IP} \ [_{NP} \ \text{John}_i \ ] \ [_{I'} \ [_{I} \ +tense \ ] \ [_{VP} \ [_{NP} \ e_i \ ] \ [_{V'} \ [_{V} \ \text{believes} \ ] \ [_{NP} \ h \ ]]]]]$
$\quad [_{IP} \ [_{NP} \ \text{John}_i \ ] \ [_{I'} \ [_{I} \ +tense \ ] \ [_{VP} \ [_{NP} \ e_i \ ] \ [_{V'} \ [_{V} \ \text{believes} \ ] \ [_{CP} \ h \ ]]]]]$
$\quad [_{IP} \ [_{NP} \ \text{John}_i \ ] \ [_{I'} \ [_{I} \ +tense \ ] \ [_{VP} \ [_{NP} \ e_i \ ] \ [_{V'} \ [_{V} \ \text{believes} \ ] \ [_{IP} \ h \ ]]]]] \ \} \ )$

The determiner *the* is now input into the buffer:

(204)

Buffer:
$\{ \ [_{NP} \ [_{DetP} \ [_{Det'} \ [_{Det} \ \text{the} \ ]]] \ [_{N'} \ [_{N} \ h \ ]]] \ \}$

The H-node NP in the buffer can attach in three positions: 1) as the direct object of the verb *believes*; 2) as subject of the IP complement; or 3) as subject of the IP argument of the CP complement of *believes*. Furthermore, an H-node genitive NP is pushed onto a stack since this H-node NP does not attach anywhere. Thus there are now two stacks in the current stack-set:

(205)

Stack-set:
$( \ \{ \ [_{IP} \ [_{NP} \ \text{John}_i \ ] \ [_{VP} \ [_{NP} \ e_i \ ] \ [_{V'} \ [_{V} \ \text{believes} \ ] \ [_{NP} \ [_{DetP} \ \text{the}] \ h \ ]]]]$
$\quad [_{IP} \ [_{NP} \ \text{John}_i \ ] \ [_{VP} \ [_{NP} \ e_i \ ] \ [_{V'} \ [_{V} \ \text{believes} \ ] \ [_{CP} \ [_{C} \ e \ ] \ [_{IP} \ [_{NP} \ [_{DetP} \ \text{the} \ ] \ h \ ] \ h \ ]]]]]$
$\quad [_{IP} \ [_{NP} \ \text{John}_i \ ] \ [_{VP} \ [_{NP} \ e_i \ ] \ [_{V'} \ [_{V} \ \text{believes} \ ] \ [_{IP} \ [_{NP} \ [_{DetP} \ \text{the} \ ] \ h \ ] \ h \ ]]]] \ \} \ )$
$( \ \{ \ [_{NP} \ [_{DetP} \ \text{the} \ ] \ [_{N'} \ [_{N} \ h \ ]]] \ \}$
$\quad \{ \ [_{IP} \ [_{NP} \ \text{John}_i \ ] \ [_{VP} \ [_{NP} \ e_i \ ] \ [_{V'} \ [_{V} \ \text{believes} \ ] \ [_{NP} \ h \ ]]]]$
$\quad [_{IP} \ [_{NP} \ \text{John}_i \ ] \ [_{VP} \ [_{NP} \ e_i \ ] \ [_{V'} \ [_{V} \ \text{believes} \ ] \ [_{CP} \ h \ ]]]]$
$\quad [_{IP} \ [_{NP} \ \text{John}_i \ ] \ [_{VP} \ [_{NP} \ e_i \ ] \ [_{V'} \ [_{V} \ \text{believes} \ ] \ [_{IP} \ h \ ]]]] \ \} \ )$

Structures for the word *light* now enter the buffer. The word *light* is ambiguous among adjective, nominal and verbal readings. Thus an adjectival phrase dominated by an H-node NP, an NP dominated by an H-node IP, a VP and an IP all now occupy the buffer:

---

[77]For simplicity, I will omit hypothesized adjunct $\overline{\text{X}}$ levels in this example, since none of these ever gets matched.

(206)

Buffer:

$\{\ [_{NP}\ [_{N'}\ [_{AP}\ [_{A'}\ [_{A}\ \text{light}\ ]]]\ [_{N'}\ [_{N}\ h\ ]]]]$

$[_{IP}\ [_{NP}\ [_{N'}\ [_{N'}\ [_{N}\ \text{light}\ ]]\ (\text{*or*}\ [_{PP}\ h]\ [_{CP}\ h\ ])\ ]]\ [_{I}\ h\ ]\ ]]]$

$[_{VP}\ [_{NP}\ e\ ]\ [_{V'}\ [_{V'}\ [_{V}\ \text{light}\ ]\ [_{NP}\ h\ ]]\ (\text{*or*}\ [_{PP}\ h\ ]\ [_{CP}\ h\ ])\ ]]$

$[_{IP}\ [_{NP}\ e\ ]\ [_{I'}\ [_{I}\ +tense\ ]\ [_{VP}\ [_{NP}\ e\ ]\ [_{V'}\ [_{V'}\ [_{V}\ \text{light}\ ]\ [_{NP}\ h\ ]]\ (\text{*or*}\ [_{PP}\ h\ ]\ [_{CP}\ h\ ])\ ]]]]\ \}$

The adjectival and nominal readings of *light* can attach to each of the H-node NPs in the structures on top of each stack. Since there are three subcategorizations for the verb *believe* and two compatible categorizations of the word *light*, there are therefore six possible attachments. One of these includes the attachment of the NP *the light* as direct object of the verb *believes*. Note that since semantic and pragmatic plausibility of syntactic structures is not checked in the syntactic parser, this parse is just as acceptable as the other syntactic parses.

The verbal entries for *light* cannot attach to either stack. Since neither of these representations contains a root H-node, both are pruned from the parse by the Stack Push Constraint. Thus although all readings of a word are accessed by the proposed parsing model, many readings die away almost immediatedly because their inability to attach or be part of a single cell representation for the input.

The word *man* is processed next, followed by the genitive marker *'s*. The NP H-nodes on the single cell stack are not compatible with genitive case, so the genitive marker *'s* can attach only to an NP H-node on top of the two cell stack. The resulting genitive NP now attaches in the argument positions of each reading of *believes*. Only the genitive NP H-node cannot attach, and so this reading is saved on top of a new stack. Thus there are still two stacks in the stack-set after the processing of the input *John believes the light man's*:

(207)

Stack-set:

$(\ \ \{\ \ [_{IP}\ [_{NP}\ \text{John}_i\ ]\ [_{VP}\ [_{V'}\ [_{V}\ \text{believes}\ ]\ [_{NP}\ [_{NP}\ \text{the light man 's}\ ]\ h\ ]]]]]$

$[_{IP}\ [_{NP}\ \text{John}_i\ ]\ [_{VP}\ [_{V'}\ [_{V}\ \text{believes}\ ]\ [_{CP}\ [_{C}\ e\ ]\ [_{IP}\ [_{NP}\ [_{NP}\ \text{the light man 's}\ ]\ h\ ]\ h\ ]]]]\ ]]]$

$[_{IP}\ [_{NP}\ \text{John}_i\ ]\ [_{VP}\ [_{V'}\ [_{V}\ \text{believes}\ ]\ [_{IP}\ [_{NP}\ [_{NP}\ \text{the light man 's}\ ]\ h\ ]\ h\ ]]\ ]]]\ \}\ )$

$(\ \ \{\ \ [_{NP}\ [_{NP}\ \text{the light man 's}\ ]\ [_{N'}\ [_{N}\ h\ ]]]\ \}$

$\{\ \ [_{IP}\ [_{NP}\ \text{John}_i\ ]\ [_{VP}\ [_{V'}\ [_{V}\ \text{believes}\ ]\ [_{NP}\ h\ ]]]]$

$[_{IP}\ [_{NP}\ \text{John}_i\ ]\ [_{VP}\ [_{V'}\ [_{V}\ \text{believes}\ ]\ [_{CP}\ h\ ]]]]$

$[_{IP}\ [_{NP}\ \text{John}_i\ ]\ [_{VP}\ [_{V'}\ [_{V}\ \text{believes}\ ]\ [_{IP}\ h\ ]]]]\ \}\ )$

This parse state is very similar to that depicted in (205) after the input *John believes the* has been processed.

The word *friend* is input next. As before with the nominal readings of *light* and *man*, the nominal reading of *friend* attaches to each NP H-node on top of each stack.

The word *ate* now enters the buffer as a tensed IP. This structure matches the hypothesized IP below the CP node subcategorized by one reading of *believes*. However, no other attachments are possible. Since the tensed IP is not an H-node, it cannot be pushed onto any stack due to the Stack Push Constraint. As a result there is only a single stack which contains the following structure after the word *ate* is processed:

(208) $[_{IP}$ $[_{NP}$ John$_i$ ] $[_{VP}$ $[_V$ believes ] $[_{CP}$ $[_{IP}$ $[_{NP}$ the light man 's friend ] $[_{VP}$ $[_V$ ate ] $[_{NP}$ *h* ]]]]]]

The words *the squid* are then processed as the object of the verb *ate*, resulting in the following final parse for (202):

(209) $[_{IP}$ $[_{NP}$ John$_i$ ] $[_{VP}$ $[_V$ believes ] $[_{CP}$ $[_{IP}$ $[_{NP}$ the light man 's friend ] $[_{VP}$ $[_V$ ate ] $[_{NP}$ the squid ]]]]]]

### 4.4.6. Reanalysis

It is assumed that reanalysis (backtracking) takes place as soon as breakdown as defined in the Single Cell Hypothesis occurs.[78] There exist many theoretically possible reanalysis algorithms. One possible algorithm consists of going back to the beginning of the input string and starting the parse over, consciously searching for possible ambiguities. This hypothesis, labelled the *forward reanalysis hypothesis* turns out to be disconfirmed by data collected by Frazier & Rayner (1982). A second plausible reanalysis algorithm consists of backing up from the point of failure one word at a time while searching for possible ambiguities that might lead to alternative analyses. This hypothesis, introduced by Kaplan (1972), is labelled the *backward reanalysis hypothesis* by Frazier & Rayner (1982). A third possible reanalysis algorithm marks locations associated with ambiguity and returns to the most recent of these sites when processing breakdown has occurred. This hypothesis was proposed by Frazier & Rayner (1982) following Winograd (1972) and Carpenter & Daneman (1981) and is called the *selective reanalysis hypothesis.*

Frazier & Rayner (1982) performed experiments that led them to conclude that some variant of the selective reanalysis hypothesis is the reanalysis algorithm employed by the human parser. However, their conclusions crucially rest on the assumption that the human parser operates serially. Since this assumption is not necessarily correct (see the discussion in Chapters 1 and 2), the conclusions do not necessarily hold. While the experiments performed by Frazier & Rayner (1982) convincingly rule out the forward reanalysis hypothesis, these experiments do not rule out a backward reanalysis algorithm operating under a parallel processing architecture. However, no work has been done since then which successfully teases apart these two possibilities.

As a result of the lack of relevant data, I will simply assume a particular reanalysis algorithm, subject to future (dis)confirmation. The algorithm that I will assume is a variant of the selective reanalysis hypothesis. Recall that in Chapter 1 I hypothesize that a structure

[78]See Chapters 1 and 6 for my assumptions with respect to what reanalysis is psychologically costly (conscious).

is *pruned* if there exists another coexisting structure of significantly smaller processing cost, where cost is measured in Processing Load Units (PLUs). This hypothesis makes up my assumption regarding local parsing preferences. I now propose a variant of this hypothesis in which the expensive structure is not actually *pruned*: rather it is set aside so that no further work will be done on it except in circumstances of reanalysis.[79] Given the existence of local parse failures, the reanalysis algorithm that I propose consists of repeating the step given in (210) until it succeeds or until it can apply no more because there are no further words to reanalyze:

(210) Reanalysis: Given no single cell representation for the current input string, find the structure(s) that were set aside most recently because of local preference constraints. Parse left to right from the word associated with these structures as defined by the underlying parsing algorithm, but without local preference constraints on reanalysis at this state. A parse failure occurs if reanalysis does not lead to a successful parse.

As observed above, this simple reanalysis procedure may not be correct. Many variants of selective and backward reanalysis are possible. For example, it may be that the parser first attempts structures for which the local difference between the previously attempted structures and the saved structures is at a minimum. Until there is empirical data that can distinguish among the many possible algorithms, I will assume the reanalysis algorithm in (210).

[79]In future chapters it is stil sometimes said that a structure is *pruned* because of its excess cost. Consider this a shorthand for saying that the expensive structure is set aside, not to be worked on again except in reanalysis situations.

## 5. Dynamic Application of the $\theta$-Criterion and Projection Principle

Following much current work in syntactic theory, I assume that an important component of the grammar of a natural language is a set of constraints, each of which rules out representations lacking a necessary property (see, for example, Chomsky (1981, 1986a, 1986b), Kaplan & Bresnan (1982), Pollard & Sag (1987)). A constraint-based syntactic theory presents an obvious set of candidates for load-bearing structural properties: the set of local violations of the filters. That is, given a constraint-based syntactic theory, it is reasonable to assume that there is a processing weight associated with the local violation of some of the syntactic filters. In particular, I will consider the $\theta$-Criterion and Projection Principle from Government and Binding Theory (Chomsky (1981, 1986a, 1986b)) with respect to the theory of processing.[80,81] The Projection Principle is given in (213):

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

GB Theory assumes the existence of a number of levels of representation. I assume that the level most relevant to parsing is surface structure (S-structure). Thus the Projection Principle applied to S-structure dictates that lexical requirements be satisfied at that level. The $\theta$-Criterion is given in (214):

(214) The $\theta$-Criterion:
Each argument bears one and only one $\theta$-role (thematic role) and each $\theta$-role is assigned to one and only one argument (Chomsky (1981) p. 36).

Note that the second part of the $\theta$-Criterion – that each $\theta$-role be assigned – is not identical to the Projection Principle applied at S-structure since not all lexical requirements involve thematic roles. For example, a complementizer subcategorizes for an IP, but assigns it no thematic role. Since, however, all thematic requirements are lexical requirements, the second part of the $\theta$-Criterion follows from the Projection Principle. Thus the $\theta$-Criterion

---

[80] As was pointed out in footnote 14, any syntactic theory which contains correlates of the relevant parts of the $\theta$-Criterion and Projection Principle will also derive the properties proposed here. Thus the Completeness and Coherence Conditions (Kaplan & Bresnan (1982)) which constrain possible f-structures in Lexical Functional Grammar (Bresnan (1982)) would give similar results:

(211) Completeness:
An f-structure is *locally complete* if and only if it contains all the governable grammatical functions that its predicate governs. An f-structure is *complete* if and only if it and all its subsidiary f-structures are locally complete. (Kaplan & Bresnan (1982), pp. 211-212)

(212) Coherence:
An f-structure is *locally coherent* if and only if all the governable grammatical functions that it contains are governed by a local predicate. An f-structure is *coherent* if and only if it and all its subsidiary f-structures are locally coherent. (Kaplan & Bresnan (1982), pp. 212)

[81] The local violation of other syntactic principles may also be associated with processing weight. I leave it to future work to determine the effects of other such local violations.

that I will assume consists only of the first clause of (214) (*cf.* Gibson (1990a, 1990b).):[82]

(215) The $\theta$-Criterion (simplified): Each argument bears one and only one $\theta$-role.

The Projection Principle rules out S-structure representations which contain lexical entries whose lexical requirements are not satisfied in those representations, where the lexical requirements of a lexical entry are the subcategorized elements of that entry. For example, the Projection Principle rules out the structures in (216):[83]

(216) a. * $[_{IP}$ $[_{NP}$ Mary ] $[_{VP}$ $[_{V'}$ $[_{V}$ found ]]]]

b. * $[_{IP}$ $[_{NP}$ John ] $[_{VP}$ thinks $[_{CP}$ $[_{C'}$ $[_{C}$ that ]]]]]

The verb *found* requires a noun phrase direct object. However, in (216a), these lexical requirements are currently unsatisfied. Thus the Projection Principle rules out (216a). Similarly, the complementizer *that* requires a following IP. None appears in (216b) and hence this structure is ruled out.

The simplified $\theta$-Criterion as given in (215) rules out structures in which a constituent requiring a thematic role receives no thematic role. For example the structures in (217) all violate the $\theta$-Criterion:

(217) a. * $[_{IP}$ $[_{NP}$ John ] $[_{VP}$ $[_{V'}$ $[_{V}$ slept ] $[_{NP}$ Mary ]]]]

b. * $[_{IP}$ $[_{NP}$ Bill ] $[_{NP}$ Fred ] $[_{VP}$ likes $[_{NP}$ Sue ]]]

c. * $[_{CP}$ $[_{NP}$ who ] $[_{IP}$ does $[_{NP}$ Fred ] $[_{VP}$ like $[_{NP}$ Sue ]]]

In (217a) the NP *Mary* is in an argument position but receives no thematic role, since the verb *slept* assigns its only thematic role to its subject, *John*. Similarly, the NPs *Bill* and *who* receive no thematic roles in (217b) and (217c) respectively. The only $\theta$-role assigner in each, the verb *like*, assigns its thematic roles to the NPs *Fred* and *Sue*.

## 5.1. The Property of Thematic Reception

The dynamically applied $\theta$-Criterion can be stated as the following processing property:

[82]The biuniqueness condition on $\theta$-assignment in the statement of the $\theta$-Criterion might be problematic (see Jackendoff (1987) and the references cited there). The key idea behind the $\theta$-Criterion is to block arguments from acquiring thematic roles at different points in the derivation of a structure. As a result, Chomsky (1981, p. 336) gives an alternative definition of the $\theta$-Criterion in terms of chain positions. None of the work here appeals appeals to the biuniqueness of thematic role assignment. The important part of the $\theta$-Criterion for this work is that arguments require thematic roles and that lexical requirements be satisfied.

[83]As noted in Chapter 4, I will be following the category notation in Chomsky (1986b) as follows: 1) the category complementizer phrase (CP) is the equivalent of the traditional $S'$ node; and 2) the category IP (Infl phrase) is the equivalent of the traditional $S$ node.

(218) The Property of Thematic Reception (PTR):
Associate a load of $x_{TR}$ PLUs to each C-node constituent that is in a position that can receive a thematic role in some co-existing structure, but whose $\theta$-assigner is not unambiguously identifiable in the structure in question.

Thus the Property of Thematic Reception assigns a relative weight to a structure if an identical copy of a constituent inside that structure appears in a thematic position in some co-existing (possibly the same) structure. In order for two constituents to be considered identical they must represent the same segment of the input string and they must have exactly the same $\overline{\mathrm{X}}$ structure, including the same lexical items as structural heads.

## 5.2. The Property of Lexical Requirement

The dynamically applied Projection Principle gives a property similar to the PTR. This property is stated in terms of *thematic* elements. Following recent work in linguistic theory, I distinguish two kinds of categories: *thematic* categories (also known as *content* or *substantive* categories) and *functional* categories (see, for example, Fukui and Speas (1986) and Abney (1987) and the references cited in each). There are a number of properties that distinguish functional elements from thematic elements, the most crucial being that functional elements mark grammatical or relational features while thematic elements pick out a class of objects or events. Thematic categories include nouns, verbs, adjectives and many prepositions; functional categories include determiners, complementizers, inflection markers and nonpredicative instances of prepositions like *of*. I hypothesize that the presence of a C-node thematic element in the subcategorization pattern of a lexical entry is sufficient for the satisfaction of that lexical requirement. The Property of Lexical Requirement, the dynamic version of the Projection Principle applied at S-structure, is given in (219):

(219) The Property of Lexical Requirement (PLR):
Associate a load of $x_{LR}$ PLUs to each lexical requirement that is obligatory in some co-existing structure, but is satisfied by an H-node constituent containing no thematic elements in the structure in question.

While all lexical requirements (subcategorization patterns) are predicted when a word is input, only those that can appear to the *right* of the head cause the construction of H-nodes. Thus only those lexical requirements that are needed to the right of a given head are associated with load via the Property of Lexical Requirement.

Also note that since all rightward lexical requirements minimally involve the existence of a hypothesized structure, the Property of Lexical Requirement ignores those structures whose lexical requirements are satisfied by either confirmed node or hypothesized nodes containing thematic elements. The PLR will therefore penalize only those structures with unsatisfied lexical requirements, where unsatisfied requirements consist of thematic element-less hypothesized structures.

Consider the parse of (220) with respect to the PLR and PTR:

(220) Fred likes Sue.

When the NP *Fred* is initially input, one structure will contain this NP as specifier of a matrix clause, a position that can and eventually will receive a thematic role. Hence this structure is associated with $x_{TR}$ PLUs. When the verb *likes* is then input and attached as matrix verb, the processing load associated with the NP *Fred* disappears, since a thematic role is now assigned to this NP. However, this structure is associated with a $x_{LR}$ PLUs since the verb *likes* requires a noun phrase complement, which is not yet satisfied. Finally the NP *Sue* is input and attached as the object of the verb *likes*, thus satisfying the lexical requirements of this verb. Since 1) all arguments receive thematic roles in the resulting structure and 2) all obligatory lexical requirements are satisfied in this structure, this structure is associated with no load via either of the Properties of Thematic Reception or Lexical Requirement.

As a result the Properties of Lexical Requirement and Thematic Reception, structures resulting from argument attachment are associated with less processing weight than are structures resulting from adjunct attachment. Consider, for example, a phrase that has two possible attachments: as an adjunct or as an argument. The first attachment will be associated with load via both the PTR and the PLR, since 1) a phrase is attached in non-thematic position, when a thematic attachment site is available; and 2) a lexical requirement of a head are satisfied in the structure resulting from argument attachment, while the same lexical requirement is not satisfied in the structure resulting from adjunct attachment. Neither of these loads will be associated with the structure resulting from argument attachment.

## 5.3. The Property of Thematic Transmission

An additional factor to be considered when examining principles of interpretation is that of thematic role transmission. I assume that thematic roles seek out categories that contain thematic content. Thus I assume that a thematic role which is assigned to a semantically null category is passed on to the argument of that category. The English complementizer *that* is an example of a semantically null category, as is the English preposition *of*. Consider a simple sentence in which a complementizer phrase is in an argument position:

(221) Fred believes [$_{CP}$ that [$_{IP}$ John saw Mary ]]

In (221) the verb *believe* assigns the thematic role PROPOSITION to its complement clause *that John saw Mary*. Note that the CP headed by *that* adds nothing to the meaning of this sentence: it is present for syntactic reasons only. The logical argument of the verb *believes* is the proposition corresponding to the IP *John saw Mary*: the complementizer phrase mediates between the verb *believes* and this IP. As a result, I assume that categories lacking semantic content, like the complementizer *that*, transmit the thematic roles that they receive to their complements. In (221), for example, I assume that the thematic role PROPOSITION received by the CP *that John saw Mary* is passed on to the complement of *that*, the IP *John saw Mary*, to give the resultant thematic relations. Note that no such thematic role transmission takes place if the complementizer heads an adjunct phrase as in (222):

(222) The dog that bit me ran away.

Given the assumption that semantically null functional categories transmit the thematic roles that they receive, it is reasonable to assume that there is an additional load associated with the unsatisfied lexical requirements of these categories. That is, in addition to satisfying its own lexical requirements, a semantically null category like the complementizer *that* must also allow the transmission of a thematic role if it heads a phrase in an argument position. Local unsatisfaction of its lexical requirements results in additional processing load via the Property of Thematic Transmission:

(223) The Property of Thematic Transmission (PTT): Associate a load of $x_{TT}$ PLUs to each semantically null C-node category in a position that can receive a thematic role, but whose lexical requirement is currently satisfied by a hypothesized constituent containing no thematic elements.

As a result of the Property of Thematic Transmission, there is an additional load of $x_{TT}$ PLUs associated with the complementizer *that* in (221) until the word *John*, a thematic element, is processed and attached in its complement. Note that no such weight is associated with the processing of the complementizer *that* in (222) since in this sentence the complementizer *that* is not attached in a thematic position and thus does not transmit any thematic role.

## 5.4. The Processing Loads Associated with the Properties of Thematic Reception, Lexical Requirement and Thematic Transmission

When particular processing loads are associated with each of the Properties of Thematic Reception, Lexical Requirement and Thematic Transmission, a large number of empirical predictions will be made in the domains of processing overload and preference. Since all three properties deal with the interpretation of an utterance, it might be reasonable to assume as a default that the loads associated with these properties, $x_{TR}$ PLUs, $x_{LR}$ and $x_{TT}$, are the same. This assumption is especially appealing with respect to the Properties of Thematic Reception and Lexical Requirement since the $\theta$-Criterion and Projection Principle are believed to follow from a more general principle: that of Full Interpretation (Chomsky (1986a)). If the $\theta$-Criterion and Projection Principle do in fact follow from the principle of Full Interpretation, then the PLR and the PTR reduce to a single property: that of local uninterpretability.[84]

While the assumption that the loads associated with the Properties of Thematic Reception and Lexical Requirement are the same turns out to be consistent with all the data that I have gathered in this thesis, this assumption does not extend to the Property of Thematic Transmission: the load associated with the Property of Thematic Transmission must

[84]The principle of Full Interpretation has not yet been adequately formalized, so I will appeal only to its current components: the $\theta$-Criterion and Projection Principle.

be distinct from the loads associated with the other two properties as a result of certain empirical evidence. Thus I will assume only two independent variables in future chapters: one associated with the Properties of Thematic Reception and Lexical Requirement and one associated with the Property of Thematic Transmission. I will continue to refer to the load associated with the PTT as the variable $x_{TT}$. The variable that I will use to refer to the Properties of Thematic Reception and Lexical Requirement will be $x_{Int}$, where *Int* is short for *interpretation*.[85]

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

However, it should be kept in mind that the loads associated with the PLR and the PTR may have independent values which just happen to be close enough together that the data that I have examined so far does not distinguish them. That is, the three properties all have independent processing loads and it just happens that two of them are close together. Even if this is the case, the presentation of the data is greatly simplified by the reduction of two variables ($x_{TR}$ and $x_{LR}$) into one ($x_{Int}$). Hence in the data to follow, I will present the load associated with the Property of Thematic Transmission as one load – $x_{TT}$ – and the loads associated with the Properties of Thematic Reception and Lexical Requirement as another – $x_{Int}$.

[85] See Gibson (1990a) for work that assumes that the loads associated with these two properties are distinct.

## 6. Ambiguity and the Properties of Thematic Reception, Lexical Requirement and Thematic Transmission

This section examines locally ambiguous sentences that either cause or do not cause garden-path effects in order to determine a value for the processing loads associated with $x_{Int}$ and $x_{TT}$ in terms of the preference constant $P$ in PLUs. The judgments of processing difficulty are based on my intuitions coupled with the intuitions of many others who participated in informal acceptability judgment tasks. A sentence is judged difficult if any reanalysis is noticed; that is, when backtracking becomes conscious.

### 6.1. Noun Phrase Complement / Sentential Complement Ambiguities

Consider the following pairs of sentences with respect to the PTR and PLR ((225e) – (225j) from Frazier & Rayner (1982)):

(225) a. Bill expected Mary.
b. Bill expected Mary to like John.
c. Bill knew John.
d. Bill knew John liked Mary.
e. The city council argued the mayor's position forcefully.
f. The city council argued the mayor's position was incorrect.
g. Tom heard the gossip.
h. Tom heard the gossip wasn't true.
i. Sherlock Holmes didn't suspect the countess.
j. Sherlock Holmes didn't suspect the countess was a fraud.

Each of the matrix verbs in the sentences in (225) is ambiguous: either subcategorizing for an NP complement or taking a sentential complement (either IP or CP). For example, the verb *expect* is ambiguous: either taking an NP complement as in (225a) or taking an IP complement as in (225b). Since there is no lexical complementizer in any of these example sentences, there is a local ambiguity in each at the point of parsing the complement NP. Despite this local ambiguity, there is no difficulty parsing either the NP complement examples or the sentential complement examples. Consider the state of the parse of (225b) after the word *Mary* has been processed:

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

In (226a) the NP *Mary* is attached as the NP complement of *expected.* In this representation there is no load associated with either of the Property of Thematic Reception since all constituents that are positions to receive thematic roles, do so. Furthermore, this structure is associated with no load via the Property of Lexical Requirement since all lexical requirements are satisfied. Finally, the fact that all lexical requirements are satisfied implies that no load is associated with this representation via the Property of Thematic Transmission. In (226b) 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 Properties of Lexical Requirement or Thematic Transmission, however, since the lexical requirements of the verb *expected* are satisfied by nodes that contain thematic elements. Since there is no difficulty in processing either (225a) or (225b), the load difference between the two structures in (226) cannot be greater than $P$ PLUs, the preference factor assumed in inequality (13). Thus the inequality in (227) is obtained:

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

Since the load difference between the two structures is not sufficient to cause a strong preference, both structures are retained. Note that this is a crucial difference between the theory presented here and the theory presented in Frazier & Fodor (1978), Frazier (1979), Frazier & Rayner (1982), Pritchett (1988) and Abney (1989). In each of these theories, only one representation can be retained, so that either (226a) or (226b) would be preferred at this point.[86]

Consider now the sentences in (228) with respect to the Properties of Thematic Reception and Lexical Requirement:

(228) a. We expected her.
b. We expected her mother to like Fred.
c. The populace believed the American.
d. The populace believed the American president ate broccoli on Tuesdays.
e. The waitress saw the light.
f. The waitress saw the light beer tumble to the floor.
g. Ben knew the Canadian.
h. Ben knew the Canadian government would not be pleased.

As in the previous examples, the matrix verb in each of these sentences subcategorizes for either an NP or a sentential complement. However, the ambiguity is longer in the sentences in (228) than in (225) because of the insertion of a categorically ambiguous word before the post-verb NP. In each case this word has at least two readings: a noun reading and a prenominal modifier reading. In particular, the word *her* has accusative noun and genitive noun lexical entries and the words *American*, *light* and *Canadian* each have adjectival and nominal lexical entries. Consider the state of the parse of (228b) after the word *her* has been processed. In one representation the NP *her* will be attached as the NP complement

[86]Frazier and her colleagues would in fact argue that their experimental data support the hypothesis that only one representation is retained at this point. See, however, Section 2.2.1 for a description of counter-evidence due to Holmes, Kennedy & Murray (1987) and Kennedy *et al* (1989).

of *expected*:

(229) $[_{IP}$ $[_{NP}$ We ] $[_{VP}$ expected $[_{NP}$ her ]]]

In this representation there is no load associated with either of the Properties of Thematic Reception or Lexical Requirement since no thematic objects need thematic roles and all lexical requirements are assigned. In another representation the NP *her* is the specifier of a hypothesized NP on a substack whose lowest level contains the other reading of the verb *expected*:

(230) { { $[_{IP}$ $[_{NP}$ We ] $[_{VP}$ expected $[_{IP}$ ]]] }
{ $[_{NP}$ $[_{NP}$ her ] ] } }

This representation is associated with at least $x_{LR}$ PLUs since the verb *expected* has a currently unsatisfied lexical requirement. However, no load is associated with the genitive NP specifier *her* since its $\theta$-assigner, although not yet present, is unambiguously identified as the head of the NP to follow (Chomsky (1986a)).[87] Hence the total load associated with (230) is $x_{LR} = x_{Int}$ PLUs. Thus the load difference between structures (229) and (230) is $x_{Int}$ PLUs. Since this load difference is not enough to cause local pruning of the more expensive structure (see (227)), the lack of difficulty in parsing any of the sentences in (228) is explained.

## 6.2. Object Noun Phrase Thematic Role Ambiguity

Consider now the sentences in (233) ((233a) and (233b) from Carlson & Tanenhaus (1988)):

(233) a. I loaded the truck on the ship.
b. I loaded the truck with beans.
c. John sprinkled the sand on the road.
d. John sprinkled the sand with water.
e. Sam cleared the tables from the room.
f. Sam cleared the tables of dishes.

Each of these sentences is locally ambiguous at the point of parsing the post-verbal noun phrase. This noun phrase may attach to the matrix verb in a position that may receive one

[87] Note that specifiers do not always receive their thematic roles from the categories which they specify. For example, a non-genitive noun phrase may specify any major category. In particular, it may specify an IP or a CP. But the specifier of these categories may receive its thematic role through chain formation from a distant $\theta$-assigner, as in (231):

(231) John appears to like beans.

Note that there is no NP that corresponds to (231) (Chomsky (1970)):

(232) * John's appearance to like beans.

of two thematic roles. In (233a) and (233b) for example, the NP *the truck* may attach as the THEME of the verb *loaded*, as in (233a), or as the LOCATION of the verb *loaded*, as in (233b). Neither of these attachments is preferred over the other, as is demonstrated by the lack of difficulty experienced in parsing either (233a) or (233b). The fact that both sentences are easy to parse is easily explained by the framework proposed here. The loads corresponding to each of the possible representations for the input string *I loaded the truck* are identical, since: 1) the $\theta$-grids for each representation of *loaded* is equally full, albeit with different roles assigned in each grid; and 2) each argument is assigned a thematic role. Thus there is no processing load difference between the two representations, and both are maintained.[88]

### 6.3. Prepositional Phrase Attachment: Adjunct / Argument Ambiguity

Consider the following sentences, each of which has local prepositional phrase attachment ambiguities:[89]

(234) a. # I put the candy on the table into my mouth.
b. # The cook placed the cake in the oven onto the table.
c. # Ron gave the letter to Nancy to the postman.

In (234a) the preposition *on* 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. A similar effect results in each of the other examples in (234). 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 in the parse of (234a):

(235) 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 (235a) is $x_{Int}$ 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 (235b) is $2x_{LR} + x_{TR} = 3x_{Int}$ PLUs since 1) both the verb *put* and the preposition *on* have unsatisfied 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.[90] Thus the

[88]In fact, it may be that there is only one syntactic representation for each of these local attachments, with multiple thematic assignments associated with this structure. The approach here does not rule out such a possibility.

[89] When presented by itself, each of these sentences invokes a garden-path effect. Note that reading multiple examples of similar garden-path effects may reduce the effect in the latter cases, since the pattern can be noticed and accounted for. See also footnote 102.

[90]Note that the $\overline{\text{X}}$ structures of the argument and adjunct representations of prepositional phrases are identical. If they weren't identical, they would not be comparable by the Property of Thematic Reception.

difference between the loads associated with the two structures is $2x_{Int}$ PLUs. Since the argument attachment structure is strongly preferred over the other structure, I hypothesize that this load is greater than $P$ PLUs:

(236) $2x_{Int} > P$

A similar explanation applies to the rest of the examples in (234). Consider now (237), which contains ambiguity with respect to attachment to the verb *send*:

(237) # The president sent the note to Gorbachev to his secretary for proofreading.

The verb *sent* in (237) differs from the verbs in the sentences in (234) in that it does not obligatorily require a PP goal argument in the syntax. In spite of this fact, the explanation of the garden-path effect in (237) is very similar to the explanation of the garden-path effects in (234). Consider the state of (237) at the point of parsing the preposition *to*:

(238) a. $[_{IP}$ $[_{NP}$ The president $]$ $[_{VP}$ $[_{V'}$ $[_{V}$ sent $]$ $[_{NP}$ the note $]$ $[_{PP}$ to $[_{NP}$ $]]$ $]]]$
b. $[_{IP}$ $[_{NP}$ The president $]$ $[_{VP}$ $[_{V'}$ $[_{V}$ sent $]$ $[_{NP}$ the note $[_{PP}$ to $[_{NP}$ $]]]$ $]]]$

As in the previous examples, the load associated with (238a) is $x_{Int}$ PLUs since the preposition *to* has an unsatisfied lexical requirement. Furthermore, the load associated with (238b) is $3x_{Int}$ PLUs since: 1) the PP headed by *to* receives a thematic role in the argument attachment structure, while it receives no such role in the current structure; 2) the preposition *to* has an unsatisfied lexical requirement; and 3) the verb *sent* has a lexical requirement corresponding to the goal PP argument which is obligatory in another structure – the argument attachment structure – which is unsatisfied in the current structure. The load difference between the two structures is therefore $2x_{Int}$ and the garden-path effect is predicted, as desired.[91,92]

---

[91]Following Larson (1988) and the references cited there, I assume that each of the prepositions in (234) assigns a thematic role to its NP complement, so that none of these prepositions is semantically null. In particular, the prepositions *in* and *on* specify locations, while the preposition *to* specifies a goal of motion along some path. While the semantic content of the preposition *to* is redundant given the presence of a verb like *send* or *give*, the preposition *to* still has semantic content: it just happens that this content is duplicated in its governing verb (Larson (1988)). Since these prepositions are all semantically contentful, the Property of Thematic Transmission does not apply to them.

[92]The explanation for the garden-path effects in (234) and (237) crucially relies on the assumption that the prepositional phrases in question are adjuncts when attached to noun phrases. While this is certainly true in the case of (234a) and (234b), it is not as clear with respect to (234c) and (237). Determining whether to classify a phrase as an obligatory argument, an optional argument or an adjunct of another phrase is a difficult problem, since different tests can give different results (see Pollard & Sag (1987) for a list of possible tests). For example, extraction from an adjunct phrase is usually ungrammatical, while extraction from an argument is usually grammatical, other factors being equal (see Huang (1982) and the references cited there):

(239) a. What did you see a picture of?
b. Who did you read a book about?

(240) a. * What did you see a tree beside?
b. * Which market did you eat an apple from?

Consider now the sentences in (242):[93]

(242) a. The minister warned the president of the danger.
b. The minister warned the president of the republic of the danger.
c. The lawyer informed the chairman of the possible lawsuit.
d. The lawyer informed the chairman of the committee of the possible lawsuit.

The first occurrence of the preposition *of* in each of the sentences in (242) has two possible attachment sites: either as argument of the matrix verb, or as argument of the preceding noun. Thus, unlike the previous examples, both ambiguities involve argument attachment sites. For example, consider (242a). The preposition *of* may attach either as argument of the verb *warned* or as argument of the noun *president*. First, consider the load associated with the matrix verb attachment. This structure is associated with $x_{Int} + x_{TT}$ PLUs since 1) the preposition *of* has an unsatisfied lexical requirement; and 2) this lexical requirement of the semantically null argument preposition *of* as yet contain no thematic elements. The load associated with the attachment as argument of the NP *the president* is $2x_{Int} + x_{TT}$ PLUs since 1) the preposition *of* has an unsatisfied lexical requirement; 2) this lexical requirement of the semantically null argument preposition *of* as yet contain no thematic elements; and 3) the verb *warned* has a lexical requirement which is satisfied in another representation, but is unsatisfied in the current representation. Thus the load difference between the two structures is $x_{Int}$ PLUs, and thus no processing difficulty is predicted with either possible attachment. This result is as desired, since none of the sentences in (242) lead to processing difficulty.

See Section 9 for some discussion of attachment ambiguities involving multiple PP adjunct sites.

### 6.4. Second Object Noun Phrase / Relative Clause Ambiguity

Consider the garden-path sentences in (243):

(243) a. # John gave the boy the dog bit a dollar.
b. # I sent the child the bookcase fell on a book.
c. # Dan bought the dog the ball hit a steak.

---

Consider the following sentences with respect to those in (239) and (240):

(241) a. ? Who did Ron see a letter to?
b. ? Who did you find a gift to?
c. ? Who did the postman pick up letters to?

While the sentences in (241) are not as bad as those in (240), they are worse than those in (239). Since they are not completely grammatical, I assume that the preposition *to* marks an adjunct in these cases. However, other tests may indicate that *to* marks an optional argument. I leave it to future work to decide this complex issue. See Abney (1989) for related comments and a different approach.

[93] Thanks to Carl Pollard for pointing out examples like these to me.

These sentences are locally ambiguous at the point of parsing the second post-verbal noun phrase in each. This NP can either attach as an argument of the matrix verb, or it can attach as the subject of a relative clause modifier of the preceding NP. The argument attachment is strongly preferred, and a garden-path effect results in each of the sentences in (243), since it is the relative clause attachment that is required for a successful parse. This garden-path effect can be explained in the same way as previous examples. Consider the state of the parse of (243a) after the NP *the dog* has been processed:

(244) a. $[_{IP}$ $[_{NP}$ John ] $[_{VP}$ $[_{V'}$ $[_{V}$ gave ] $[_{NP}$ the boy ] $[_{NP}$ the dog ]]]]
b. $[_{IP}$ $[_{NP}$ John ] $[_{VP}$ $[_{V'}$ $[_{V}$ gave ] $[_{NP}$ the $[_{N'}$ $[_{N'}$ boy$_i$ ] $[_{CP}$ $[_{NP}$ $O_i$ ] $[_{IP}$ $[_{NP}$ the dog ] ]]]] $[_{NP}$ ] ]]]

While structure (244a) requires no load at this stage, structure (244b) requires $3x_{Int}$ PLUs since 1) one thematic role associated with the verb *gave* has not yet been assigned; 2) the operator in the specifier position of the CP modifying *boy* is not linked to a thematic role; and 3) the NP *the dog* is in a thematic position but does not yet receive a thematic role. Thus structure (244a) is strongly preferred and a garden-path effect results when this preference turns out to be wrong.

## 6.5. Matrix Clause / Reduced Relative Clause Ambiguities

Consider the well-known class of garden-path sentences in (245) ((245a) from Bever (1970)):

(245) a. # The horse raced past the barn fell.
b. # The boat floated down the river sank.
c. # The dog walked to the park chewed the bone.

Each of these sentences is ambiguous at the point of parsing the first verb in the sentence. The form of each of these verbs is ambiguous between simple past tense and passive participle. The simple past tense form may attach as the matrix verb of the whole sentence, while the passive participle form may attach as a reduced relative clause modifying the initial noun phrase. The matrix clause reading is strongly preferred and a garden-path effect results, since the reduced relative clause structure is the one that is needed for a grammatical parse of each sentence. To see how the garden-path effect can be accounted for in the proposed framework, consider the parse of (245a) at the point of parsing the word *raced*:

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

Structure (246a) 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. Consider now structure (246b). In (246b), the passive participle reading of *raced* has attached as a modifier of the NP *the horse*, but chain formation for the operator $O_i$ has not yet taken place. That is, although the operator $O_i$

needs to be associated with a thematic role, the search for an appropriate position from which to receive a thematic role has not yet occurred (see Section 4.4.3).[94] The load associated with structure (246b) is $2x_{Int}$ 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 can receive a thematic role but currently does not receive one. Note that the passive participle *raced* does not have any lexical requirements to the right, so no load is associated with this structure via the Property of Lexical Requirement. Thus the difference between the processing loads associated with structures (246a) and (246b) is $2x_{Int}$ PLUs. As has been seen previously, this load difference is enough for a strong local preference to take place, and a garden-path effect results.

A surprising effect occurs when a verb which *optionally* subcategorizes for a direct object, like *race*, is replaced by a verb which *obligatorily* subcategorizes for a direct object, like *find* (Pritchett (1988)):

(247) a. The bird found in the room was dead.
b. The alien examined under a microscope originated from a planet orbiting Polaris.
c. The monkeys chased out of the cage never returned.

Although the structures and local ambiguities in the sentences in (247) and (245) are similar, those in (245) cause a garden-path effect while, surprisingly, (247) do not. To determine why the sentences in (247) do not induce garden-path effects we need to examine the local ambiguity when the ambiguous verb forms in (247) are read. Consider (247a) at the point of parsing the word *found* for example:

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

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 (248a), this representation is associated with $x_{Int}$ PLUs of processing load. Like structure (246b), structure (248b) requires $2x_{Int}$ PLUs. The difference between the processing loads of structures (248a) and (248b) is therefore $x_{Int}$ PLUs. Since it was demonstrated earlier that this load difference was not sufficient to force pruning, the lack of garden-path effects in the sentences (247) is explained.

Furthermore, these results correctly predict that the sentences in (249) are not garden-

[94] As soon as chain formation takes place, the operator will be linked to the trace in the VP object position of the verb *raced* via a trace in the IP subject position. (I assume that the trace in the object position of the passive verb is posited in the lexicon accessed by the processor.) Once this chain has been formed, there will no longer be a load associated with the operator $O_i$. However, given the simple attachment algorithm, attachment must take place before chain formation occurs. Hence it is reasonable to assume that either 1) processing loads are calculated only after structures are locally complete with respect to operations such as chain formation; or 2) processing loads are calculated at any time after attachment has taken place, including the point before chains have been extended as necessary. I will assume the second alternative since it gives more desirable empirical results with respect to examples like those in (245).

path sentences either:

(249) a. The bird found in the room enough debris to build a nest.
b. The alien examined under a microscope a residue from his spaceship.
c. The monkeys chased out of the cage a large group of rats.

Consider (249a), for example. At the point of processing the word *found*, both representations in (248) are retained since the load difference between these two structures is low. Next the preposition *in* is input. This preposition attaches as a modifier of the verb *found* in both structures. In addition, a gap is placed in the direct object position of the main verb reading of *found* in order to allow the possibility of eventually satisfying its lexical requirements. However, this empty category has no thematic content and therefore does not yet satisfy the verb's lexical requirements. Thus the load difference between the two possible structures remains unchanged when only the PLR and PTR are consulted.[95] Similarly the load difference between the two possible structures does not change when the object of the preposition *the room* is processed. Eventually the NP *enough debris to build a nest* is input and attached as an adjunct bound to the direct object position of *found* in the matrix clause structure.

A final class of matrix clause/reduced relative clause ambiguities is arrived at by considering verbs that take three arguments: two internal arguments and one external one. Consider the following examples (Rayner, Carlson & Frazier (1983)):

(250) a. # The performer sent the flowers was pleased.
b. # The man lent the book never returned it.
c. # The woman brought the flowers smiled broadly.

Each of these sentences induces a garden-path effect. In order to explain this effect, consider the state of the parse of (250a) at the point of parsing the verb *sent*:

(251) a. $[_{IP}$ $[_{NP}$ the performer $]$ $[_{VP}$ sent $]]$
b. $[_{IP}$ $[_{NP}$ the $[_{N'}$ $[_{N'}$ performer$_i$ $]$ $[_{CP}$ $O_i$ $[_{IP}$ $[_{VP}$ sent $e$ $]]]$ $]]$ $]$

Suppose that the verb *sent* in structure (251a) is the non-dative shift reading of *sent*. The load associated with (251a) is then $x_{Int}$ PLUs, since only the NP theme lexical requirement is obligatorily required.[96] Now consider structure (251b), in which it is assumed that the verb *sent* is the passive form of the dative-shifted reading, so that the VP subject of *sent*

[95]There is probably additional load associated with the matrix verb structure since this structure necessitates Heavy NP Shift, a fairly marked construction in English. The slight difficulty associated with the processing of the sentences in (249) may be attributed to this additional weight.

[96]Note that the GOAL thematic role is not syntactically obligatory in these verbs, even for a verb like *send*:

(252) a. I sent the flowers.
b. The library lent the book.
c. The woman brought the flowers.

receives the goal thematic role.[97] The load associated with this structure is $3x_{Int}$ PLUs since: 1) the NP *the performer* is in a thematic position but currently lacks a thematic role; 2) the operator $O_i$ is in a thematic position but currently lacks a thematic role; and 3) the verb *send* requires that the THEME thematic role be filled, but it is currently unfilled. Thus the load difference between structures (251a) and (251b) is $2x_{Int}$ PLUs, and a strong preference for structure (251a) results. Hence the garden-path effect in sentences such as (250a) is explained.

Suppose that the passive form of the verb *sent* in (251b) is the passive of the non-dative shifted reading of *send*, so that the trace in the subject of the passive VP is associated with the THEME thematic role. The processing load associated with this structure is only $2x_{Int}$ PLUs. The load calculation for this structure is the same for that of (251b), except that no load is required by the Property of Lexical Requirement, since the verb *sent* only optionally requires a GOAL thematic role. Thus the load difference between a structure like (251b) whose trace is in direct object position and (251a) is only $x_{Int}$ PLUs, not enough to induce a garden-path effect. This is the correct prediction, as is made clear by the lack of garden-path effect in the examples in (254):

(254)a. The performer sent to the theatre was pleased.
b. The secretary lent to the department enjoyed the temporary work.
c. The dog brought for Mary ate the steak.

## 6.6. Verbal Complement Clause / Relative Clause Ambiguities

Consider the sentences in (255) and (256):

(255)a. John told the man that kissed Mary that Bill saw Phil.
b. The patient persuaded the doctor that was having trouble to leave.
c. Dan convinced the child that was frightened of dogs that cats are good pets.

(256)a. John told the man that Mary kissed Bill.
b. The patient persuaded the doctor that he was having trouble.
c. Dan convinced the child that cats are good pets.

---

Compare these sentences with ones in which the THEME thematic role is missing:

(253)a. * I sent.
b. * The library lent.
c. * The woman brought.

[97] See below for an explanation of the situation in which the verb *sent* in a structure like (251b) is the passive of the non-dative shifted reading of *send*, so that the VP subject of *sent* receives the goal thematic role.

Each of these sentences is ambiguous at the point of processing the complementizer *that*. This complementizer may attach as an argument of the matrix verb, as in the sentences in (256), or as a relative clause modifier of its preceding noun phrase, as in the sentences in (255). Neither of these attachments causes a strong preference by itself, as is demonstrated by the lack of difficulty that people have with any of these examples. In order to explain this lack of difficulty, consider the parse state in (255a) and (256a) when the complementizer *that* is processed:

(257) a. [$_{IP}$ [$_{NP}$ John ] [$_{VP}$ [$_{V'}$ [$_{V}$ told ] [$_{NP}$ the man ] [$_{CP}$ that [$_{IP}$ ]] ]] ]]]
b. [$_{IP}$ [$_{NP}$ John ] [$_{VP}$ [$_{V'}$ [$_{V}$ told ] [$_{NP}$ the [$_{N'}$ [$_{N'}$ man$_i$ ] [$_{CP}$ [$_{NP}$ $O_i$ ] that [$_{IP}$ ] ]]] [$_{CP}$ ] ]]]

The lexical requirements of the complementizer *that* are not yet satisfied in either structure, so that a load of at least $x_{Int}$ PLUs is associated with each. Structure (257a) is associated with an additional load of $x_{TT}$ PLUs since the lexical requirements of the semantically null argument complementizer *that* as yet contain no thematic elements. The total load associated with this structure is therefore $x_{Int} + x_{TT}$ PLUs. Structure (257b) is associated with a further load of $2x_{Int}$ PLUs since 1) the thematic role PROPOSITION is not yet assigned by the verb *told*;[98] and 2) the operator in the specifier position of the CP headed by *that* is not yet linked to a thematic role. Thus the total load associated with structure (257b) is $3x_{Int}$ PLUs.[99] Hence the load difference between the two structures drops to $|2x_{Int} - x_{TT}|$ PLUs. In order to explain the lack of garden-path effects in (255), it is assumed that this load difference is less than the preference constant $P$ PLUs:

(258) $|2x_{Int} - x_{TT}| \leq P$

Note that without the Property of Thematic Transmission, we would have no explanation of the lack of garden-path effects in (255). If the existence of the PTT only had consequences with respect to the class of sentences in (255), then it could be seen as stipulative. However, the PTT also has a number of empirical consequences with respect to the theory of processing overload. See Chapter 8 for a description of these effects.

While there is no difficulty in parsing the subject relative clauses in sentences like those in (255), object relative clauses are not so easily processed:

(259) a. # John told the man that Mary kissed that Bill saw Phil.
b. # The patient persuaded the doctor that he was having trouble with to leave.
c. # Dan convinced the child that the dog bit that cats are good pets.

---

[98] While this thematic role is not obligatorily assigned by the verb *tell*, it is assigned in a coexisting structure – the argument attachment structure (257a) – so that there is load associated with this lexical requirement via the Property of Lexical Requirement.

[99] It might seem that the adjunct attachment of the complementizer *that* in (257b) should be associated with an additional load of $x_{Int}$ PLUs via the PTR since the complementizer *that* is attached in an argument position in another structure, structure (257a). However, this is not so, since the CP $\overline{\text{X}}$ structures in (257a) and (257b) are not identical, and therefore are not comparable by the PTR. The CP representation for *that* in (257b) contains an operator in its specifier position, while no such operator is present in the CP representation for *that* in (257a). Thus the total load associated with structure (257b) is only $3x_{Int}$ PLUs.

That is, when a noun phrase is input following the complementizer *that* is sentences like those in (256) and (259) the attachment of that complementizer as an argument of the matrix verb is then strongly preferred. Thus the sentences in (256) are processed easily and the those in (259) induce garden-path effects. In order to explain these effects, consider the state of the parse of (256a) and (259a) after the the NP *Mary* has been input:

(260) a. $[_{IP}$ $[_{NP}$ John ] $[_{VP}$ $[_{V'}$ $[_{V}$ told ] $[_{NP}$ the man ] $[_{CP}$ that $[_{IP}$ $[_{NP}$ Mary ]]] ]] ]]]
b. $[_{IP}$ $[_{NP}$ John ] $[_{VP}$ $[_{V'}$ $[_{V}$ told ] $[_{NP}$ the $[_{N'}$ $[_{N'}$ man$_i$ ] $[_{CP}$ $[_{NP}$ $O_i$ ] that $[_{IP}$ $[_{NP}$ Mary ]] ]]] $[_{CP}$ ] ]]]

First consider structure (260a). The load associated with this structure is $x_{Int}$ PLUs since the NP *Mary* is an argument currently lacking a thematic role in this structure. Note that there is no longer any load via the Properties of Lexical Requirement or Thematic Transmission since the lexical requirements of the complementizer *that* are now satisfied in this structure by a hypothesized IP containing the thematic NP *Mary*. Now consider structure (260b). The load associated with this structure remains at $3x_{Int}$ PLUs since 1) the NP *Mary* is an argument currently lacking a thematic role; 2) the operator $O_i$ is a second NP in a position that can receive a thematic role but currently lacks one; and 3) the thematic role PROPOSITION is unassigned in the current structure. Hence the load difference between the two structures increases to $2x_{Int}$ PLUs, and the strong preference for the argument attachment is explained. Thus the garden-path effects in (259) are also explained.

Furthermore, garden-path effects are also correctly predicted if there is no lexical complementizer in complementizer phrase attachment ambiguities:

(261) a. # John told the man Mary kissed that Bill saw Phil.
b. # The patient persuaded the doctor he was having trouble with to leave.
c. # Dan convinced the child the dog bit that cats are good pets.

These effects are explained in a way exactly parallel to that for the examples in (259).

## 6.7. Nominal Complement Clause / Relative Clause Ambiguities

While a complementizer phrase prefers to be attached as a verbal argument rather than as a relative clause modifier of a preceding noun phrase, there is no strong preference when the two possible attachment sites are as argument or adjunct to the same noun. Consider the following data ((262a) and (263a) from Fodor (1985)):

(262) a. The report that the president sent to us helped us make the decision.
b. The idea that Bill came up with seemed crazy at the time.
c. The belief that Jim held kept him strong throughout his life.

(263) a. The report that the president sent the troops into combat depressed me.
b. The idea that Bill came up with the $10,000 seemed crazy at the time.
c. The belief that Jim held up the bank was persistent among many of the townspeople.

In particular, let us look at (262a) and (263a). Both sentences are initiated with the words *the report that the president sent.* In (262a), the CP *that the president sent* is attached as a relative clause modifier of the NP *the report.* In (263a), the same CP is attached as argument of the NP *the report.* Neither attachment is difficult to process. In all other cases that have been examined until this point, argument attachments have been strongly preferred over adjunct attachments, so this lack of contrast is somewhat surprising. It turns out that the ease of processing each of these constructions follows immediately from the properties and processing loads assumed thus far. Consider two possible structures for the input *the report that*:

(264) a. $[_{IP}\ [_{NP}$ the $[_{N'}\ [_{N}$ report $]\ [_{CP}$ that $[_{IP}\ ]]\ ]]]$
b. $[_{IP}\ [_{NP}$ the $[_{N'}\ [_{N'}$ report$_i$ $]\ [_{CP}\ [_{NP}\ O_i\ ]\ [_{C'}$ that $[_{IP}\ ]]]\ ]]]$

In (264a) the complementizer *that* is attached as an argument of the noun *report.* The load associated with this structure is $2x_{Int} + x_{TT}$ PLUs since 1) the NP *the report* requires a thematic role but currently lacks one; 2) the complementizer *that* has lexical requirements that are unfilled; and 3) the complementizer *that* is attached in an argument position and its lexical requirements as yet contain no thematic elements. The load associated with the relative clause attachment of the complementizer *that* as depicted in (264b) is $3x_{Int}$ PLUs since 1) the NP *the report* requires a thematic role; 2) the lexical requirements of the complementizer *that* are as yet unfilled; and 3) the operator $O_i$ in the specifier position of the relative clause requires a thematic role but currently lacks one.[100] Hence the load difference between the two structures is $|x_{Int} - x_{TT}|$ PLUs. In order to explain the lack of garden-path effect in either (262a) or (263a), I assume that this load difference is not sufficient to cause structure pruning:

(265) $|x_{Int} - x_{TT}| \leq P$

When the next words of (262a) and (263a), the NP *the president*, are input and attached, the load associated with the argument structure becomes $2x_{Int}$ PLUs since there are two NPs, *the report* and *the president*, that require thematic roles. There is no longer any load associated with the Properties of Lexical Requirement or Thematic Transmission since the lexical requirements of the complementizer *that* are now satisfied. The load associated with the relative clause attachment becomes $3x_{Int}$ since there are three NPs in this structure that require thematic roles: the lexical NPs *the report* and *the president* and the non-lexical operator $O_i$. Thus the load difference at this point becomes $x_{Int}$ PLUs, which, as we have seen earlier, is not enough to cause pruning of the more expensive structure.

When the following word of (262a) and (263a), the verb *sent*, is input and attached, the

[100]Note that there can be no load associated with the lexical requirements of the noun *report* since the argument attachment site is no longer available in structure (264b). That is, since an adjunct has already been attached to the NP *the report*, the $\overline{\text{X}}$ argument attachment site no longer exists.

load difference remains at $x_{Int}$ PLUs, although this time the relative clause reading is the less expensive of the two. As a result, both structures can be maintained, and no conscious difficulty is predicted with either, as desired.

Of course, the relative clause modifier of the head noun *report* need not be an object relative clause: it may be a subject relative clause:

(266) a. The report that motivated our investigation was submitted two days late.
b. The fact that was mentioned by John was crucial in our decision.
c. The stories that amuse John most are published by the Weekly World News.

The ease of processing these sentences is explained in the same manner. Consider, for example, (266b). The representations in (264) are retained for the initial words *the report that*. When the verbal element *was* is then input, only the relative clause structure is pursued, and no processing difficulty ensues, as desired.

### 6.8. Sentential Complement / Sentential Subject Ambiguities

Consider the sentences in (267):

(267) a. # I believe that John smokes annoys Mary.
b. # The populace knows that Dan is stupid is obvious.
c. # The students thought that Rick ate a sandwich larger than his head was impressive.

In each of the sentences in (267), the complementizer *that* can attach either as the complement of the matrix verb or as subject of the sentential complement to come. The argument attachment is strongly preferred and a garden-path effect results, since it is the other reading that is necessary for a successful parse. Consider the state of the parse of (267a) after the word *that* has been input:

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

First consider structure (268a). All positions that can receive thematic roles, do so. There is therefore no load associated with (268a) with respect to the Property of Thematic Reception. However the complementizer *that* requires an IP complement, so that $x_{Int}$ PLUs are associated with (268a). Furthermore, this complementizer is a semantically null element that will transmit the thematic role that it receives to its complement, but its complement contains no thematic elements, so it is associated with an additional load of $x_{TT}$ PLUs by the Property of Thematic Transmission. Thus the total load associated with (268a) is $x_{Int} + x_{TT}$ PLUs.

Consider now (268b). First of all, the CP headed by *that* is in a $\theta$-role receiving position, but it does not yet receive a $\theta$-role. Hence (268b) is associated with at least $x_{Int}$ 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 (268b) is associated with an additional $2x_{Int}$ PLUs via the Property of Lexical Requirement. Finally, the complementizer *that* is in a thematic position, but its lexical requirements contain no thematic elements yet, so that an additional load of $x_{TT}$ by the Property of Thematic Transmission are associated with this structure. The total load associated with structure (268b) is therefore $3x_{Int} + x_{TT}$ PLUs. The difference between the loads of the structures in (268) is therefore $2x_{Int}$ PLUs and the garden-path effects in (267) are predicted, as desired.

There exists a similar explanation of the strongly preferred readings in each of the sentences in (269):

(269) a. I believe that John smokes to annoy Mary.
b. The students believe that Rick ate a sandwich larger than his head to be impressive.

Although neither of these sentences prompts a garden-path effect, there is an ambiguity in each that goes unnoticed. The final three words in each of these sentences, initiated by the inflection marker *to*, can attach as either a purposive modifier of the preceding clause or as head of the verbal IP complement. The purposive modifier reading is very strongly preferred. The explanation for this preference is the same as for the garden-path effects in (267): when the complementizer *that* is read, the complement attachment is strongly preferred. Since the sentential subject reading is removed, the purposive modifier attachment is the only possible attachment of the inflection marker *to*. Hence the purposive modifier attachment is strongly preferred.

## 6.9. Lexical Ambiguities

### 6.9.1. Prenominal Modifier / Relative Clause Ambiguities

Consider the following class of garden-path sentences ((270b) from Marcus (1980)):

(270) a. # The Russian women loved died.
b. # The cotton clothing is made of grows in Mississippi.
c. # The computer companies buy most is the MacIntosh.

The first three words of each of these sentences can be interpreted either as a noun phrase with prenominal modifier or as a noun phrase followed by a hypothesized relative clause whose subject is present. It turns out that the simple NP reading is strongly preferred, and thus a garden-path effect results in each of the sentences in (270). This strong preference can be explained by the Properties of Thematic Reception and Lexical Requirement. In particular, let us examine the point of local ambiguity in the parse of (270a):

(271) a. $[_{IP}$ $[_{NP}$ the Russian women$]$ $]$
b. $[_{IP}$ $[_{NP}$ the $[_{N'}$ $[_{N'}$ Russian$_i$ $]$ $[_{CP}$ $[_{NP}$ $O_i$ $]$ $[_{IP}$ $[_{NP}$ women $]$ $]]$ $]]]$

Structure (271a) requires $x_{Int}$ PLUs since the NP *the Russian women* needs a thematic role but currently lacks one. Structure (271b), on the other hand, requires at least $3x_{Int}$ PLUs since 1) two noun phrases, *the Russian* and *women*, need but currently lack thematic roles; and 2) the operator in the specifier position of the modifying Comp phrase can be associated with a thematic role, but currently is not linked to one. Since the difference between these loads is $2x_{Int}$, a garden-path effect results.[101]

### 6.9.2. Adjectival / Nominal Ambiguity followed by Nominal / Verbal Ambiguity

Consider the following locally ambiguous sentences none of which causes processing difficulty (adapted from Frazier & Rayner (1987)):[102]

(273) a. The warehouse fires numerous employees each year.
b. The warehouse fires harm some employees each year.
c. The desert trains young people to be especially tough.
d. The desert trains are especially tough on young people.
e. The county buses are filled with grade school children from another county.
f. The county buses most of the grade school children to another county.

Each of the sentences in (273) consists of a determiner followed by a word which is ambiguous between adjectival and nominal readings followed by a word which is ambiguous between nominal and verbal readings. Thus there exist at least two possible structures for the first three words of each sentence: a noun phrase consisting of a determiner, and adjective and a noun; or a (partial) sentence consisting of a noun phrase followed by a verb. It turns out that neither of these readings is difficult to process, as is shown by the lack of

[101]Note that no load is associated with the adjective *Russian* since there exists no structure in which an identical adjective phrase structure is in an argument position. Furthermore, even if *Russian* in (271a) were classified as a noun as in (271b), there would be no additional load associated with this structure by the Property of Thematic Reception, since the two NP representations of *Russian* have different $\overline{\text{X}}$ structures and therefore are not comparable by the PTR. To see that the two NPs have different structures, consider the examples in (272):

(272) a. * [$_{NP}$ the [$_{NP}$ a Russian ] woman ]
b. * [$_{NP}$ a [$_{NP}$ the computer ] company ]

As is evidenced by (272), a prenominal NP modifier cannot have a specifier. In contrast, when nouns like *Russian* and *computer* are not prenominal modifiers, they must have specifiers. Thus the syntactic structures for the prenominal NP and bare NP readings of words like *computer* are syntactically distinct and are therefore not comparable by the PTR.

[102]Note that first reading a sentence that resolves ambiguity one way then reading a sentence with the same ambiguity but resolved in another way may lead to a garden-path effect because of lexical priming. That is, having just interpreted an ambiguous word one way leads one to interpret the same word in a new sentence in the same way. See also footnote 89.

garden-path effects in the sentences in (273).[103] In order to see how the model presented here explains the lack of difficulty associated with these sentences, consider the parse of sentences (273a) and (273b) at the point of parsing the word *trains*:

(274) a. $[_{IP}$ $[_{NP}$ the desert trains] ]
b. $[_{IP}$ $[_{NP}$ the desert ] $[_{VP}$ trains $[_{NP}$ ] ] ]

Structure (274a) is associated with $x_{Int}$ PLUs since it contains an NP in argument position which is not associated with a thematic role. Structure (274b) is also associated with $x_{Int}$ PLUs since the verb *trains* has a locally unsatisfied lexical requirement. The load difference between the two structures is therefore 0 PLUS, and both are maintained as desired. See also Milne (1982) and Section 9.2.

Even when the ambiguous verbal element is optionally intransitive, there is no difficulty parsing either possibility, as is exemplified by the examples in (275) (data adapted from Frazier & Rayner (1987)):

(275) a. The official lies are very transparent.
b. The official lies often.

At the point of parsing the word *lies* the load difference between the two possible structures is $x_{Int}$ PLUs, since there is no load associated with the sentential reading. However, this load difference is not enough to cause a strong preference, and both structures are retained as desired.

Similarly, when the ambiguous verbal element is obligatorily bitransitive, there is still no difficulty parsing either possibility:

(276) a. The American places that I like to visit most are in California.
b. The American places the book on the table.

The load difference between the two possible attachments is $x_{Int}$ PLUS once again, although in this case the sentential reading is associated with $2x_{Int}$ PLUs while the NP reading is still associated with $x_{Int}$ PLUs.

Furthermore, neither ambiguity is difficult to parse when the verbal element is an auxiliary verb (Milne (1986), Frazier & Rayner (1987), (277c) and (277d) adapted from Frazier & Rayner (1987)):

(277) a. The paint can be applied easily with a new brush.
b. The paint can fell down the stairs.
c. The official will make sure that no money goes to the philanthropist's son.
d. The official will did not leave any money to the philanthropist's son.

The ease of parsing the ambiguities in (277) is explained exactly as in examples (273):

[103] See Frazier & Rayner (1987) for further interesting data regarding the processing of lexical ambiguities.

both the NP and the sentential structures are associated with $x_{Int}$ PLUs at the ambiguity point, so both structures are maintained.

While the sentences in (273) are not difficult to process, Milne (1982) notes the following interesting set of examples:

(278) a. The building blocks are red.
b. The building blocks the sun.
c. # The building blocks the sun faded are red.
d. The building blocks the sun shining on the house.
e. # The building blocks the sun shining on the house faded are red.

While neither (278a) and (278b) cause processing difficulty, (278c) and (278e) cause severe garden-path effects. These garden-path effects are easily explained in the framework presented here. Consider two possible structures for the input string *the building blocks the sun*:

(279) a. $[_{IP}$ $[_{NP}$ the building ] $[_{VP}$ blocks $[_{NP}$ the sun ]]]
b. $[_{IP}$ $[_{NP}$ the building blocks $[_{CP}$ $[_{NP}$ $O_i$ ] $[_{IP}$ $[_{NP}$ the sun ] ]]]]

There is no load associated with structure (279a) since all thematic elements that require $\theta$-roles receive them and all lexical requirements are satisfied. Structure (279b), on the other hand, is associated with $3x_{Int}$ PLUs since there are three NPs requiring thematic roles but locally lacking such roles. As a result the load difference between the two structures is $3x_{Int}$ PLUs, enough to cause a strong preference for structure (279a). Thus the effects in (278) are explained.

Consider now the garden-path sentences in (280), ((280a) from Bever, 1970):

(280) a. # The editor authors the newspaper hired liked laughed.
b. # The bird flies bite most is a duck.
c. # The cotton felt is made of is inexpensive.

The three words initiating each of these sentences is ambiguous among NP readings, sentential readings and NP with relative clause modifier readings.[104] The relative clause modification is the structure required in each sentence, but this structure is locally unpreferred. Consider, for example, the parse of (280a) at the point of parsing the word *authors*:

(281) a. $[_{IP}$ $[_{NP}$ the editor ] $[_{VP}$ $[_{V'}$ $[_{V}$ authors ] $[_{NP}$ ] ]]]
b. $[_{IP}$ $[_{NP}$ the $[_{N'}$ $[_{N'}$ editor$_i$ ] $[_{CP}$ $[_{NP}$ $O_i$ ] $[_{IP}$ $[_{NP}$ authors ] ]] ]]]

The structure including the verbal reading is associated with $x_{LR}$ PLUs since the $\theta$-grid for the verb *authors* includes an unassigned role. Structure (281b), on the other hand, includes

[104]Some of the NP interpretations are semantically or pragmatically anomalous: see Section 9.3 for a description of how semantics and pragmatics interact in the proposed framework.

three noun phrases, each of which is in a position that may be linked to a thematic role but currently is not linked to any $\theta$-role. Hence the load associated with structure (281b) is $3x_{Int}$ PLUs. The load difference between these two structures is therefore $2x_{Int}$ PLUs. Thus only the cheaper of the two structures is retained, (281a), and the garden-path effect is predicted.[105]

### 6.9.3. Auxiliary / Main Verb Ambiguity

Despite the ambiguity of the *have* and *be* verbs between main and auxiliary readings in English, no difficulty is apparent in any of the following locally ambiguous sentences ((282a) and (282b) from Marcus (1980)):

(282) a. Have the boys take the exam.
b. Have the boys taken the exam?
c. Is the crowd in the room yet?
d. Is the crowd in the room happy?

Consider, for example, (282a) after the input string *have the boys* has been parsed:

(283) a. $[_{CP}$ have $[_{IP}$ $[_{NP}$ the boys ] ]]
b. $[_{IP}$ $[_{VP}$ have $[_{IP}$ $[_{NP}$ the boys ]] ]]

Structure (283a) contains the auxiliary reading of the word *have*, while (283b) contains a main verb reading. The load associated with each of these structures is $x_{Int}$ PLUs since the NP *the boys* does not receive a thematic role in either. No load is required for lexical requirements of either reading of *have* since each complement contains a thematic element, the NP *the boys*. Thus there is no load difference between the two structures, and both can be maintained in parallel. Hence the lack of garden-path effect in any of the sentences in (282) is explained.[106,107]

---

[105]Sentence (280a) is in fact more difficult to process than other garden-path sentences examined thus far. This extra difficulty is due to the fact that it contains a doubly center-embedded relative clause, whose processing independently requires more short term memory than is available. See Chapter 8 for an explanation of the unacceptability of doubly center-embedded relative clauses.

[106]Marcus (1980) notes that the noun phrase following *have* can be arbitrarily long without inducing garden-path effects one way or another. This fact was his motivation for looking ahead by syntactic constituents rather than words. See, however, Kurtzman (1985) for evidence that there is a preference for the auxiliary reading of the word *have*.

[107] While the sentences in (282) do not cause processing difficulty, (284a) and (284c) induce garden-path effects ((284a) from Marcus (1980)):

(284) a. # Have the soldiers given their medals by their sweethearts.
b. Have the soldiers given their medals to their sweethearts?
c. # Have the boys removed from the room.
d. Have the boys removed the equipment from the room?

The existence of the garden-path effects in (284a) and (284c) comes as a surprise to the approach presented

### 6.9.4. The Word *That*: Complementizer / Determiner Ambiguity: Post-Verbal Positions

The word *that* is ambiguous between at least two readings: a complementizer and a demonstrative determiner. A third reading, which is arguably derivable from the demonstrative determiner, is the demonstrative pronoun (see Jackendoff (1977) and the references cited there). Consider the sentences in (285), none of which gives the human processor any difficulty ((285b) and (285c) from Milne (1986)):

(285) a. I know that.
b. I know that boy should do it.
c. I know that boys should do it.

Given the three possible readings for the word *that*, there are at least three possible representations available when this word is processed in each of the examples in (285). Two of these structures are given as follows:

(286) a. $[_{IP}$ $[_{NP}$ I $]$ $[_{VP}$ know $[_{CP}$ that $[_{IP}$ $]]]]$
b. $[_{IP}$ $[_{NP}$ I $]$ $[_{VP}$ know $[_{NP}$ that $]]]$

Structure (286a) has the complementizer reading of *that* attached as the argument of the verb *know*. The load associated with this structure is $x_{Int} + x_{TT}$ PLUs since 1) the lexical requirements of the complementizer *that* are not yet satisfied and 2) the complementizer *that* is a semantically null functional element in an argument position that must transmit a thematic role, but does not do so yet. On the other hand, structure (286b), in which the word *that* is attached as a demonstrative pronoun, is associated with no load via the Properties of Thematic Reception, Lexical Requirement or Thematic Transmission.[108] The load difference between the two structures is therefore $x_{Int} + x_{TT}$ PLUs. Since there is no difficulty processing either (285a) or (285c), I hypothesize that this load difference is not sufficient to cause structure pruning:

(287) $x_{Int} + x_{TT} \leq P$

A third representation for *I know that* is also possible. This representation includes the determiner reading of *that* placed in a stack cell above the IP reading for *I know*. The load associated with this representation for the input *I know that* is $x_{Int}$ PLUs, since the lexical requirements of the verb *know* have not yet been satisfied. This stack representation is also retained since its processing load is not substantially greater than any others. Thus sentence (285b) can also be processed without difficulty.

---

here. While there are two possible structures for the initial substrings of the sentences in (284), the processing load differences for these structures never rises over $x_{Int}$ PLUs. Thus the garden-path effects are unexpected. See Chapter 10 for a possible explanation of these data.

[108] It may be that this structure is associated with processing load via discourse or contextual properties, since no context has been established in which the demonstrative pronoun *that* can be referential. I leave this issue to future research.

### 6.9.5. The Word *That*: Complementizer / Determiner Ambiguity: Pre-Verbal Positions

The sentences in (289) provide empirical evidence in support of the hypothesis that reanalysis is expensive if and only if that reanalysis involves previously attached words. This hypothesis, the Conscious Reanalysis Hypothesis, is repeated below:

(288) Conscious Reanalysis Hypothesis:
Reanalysis of an input string is expensive if and only if representations for some segment of that string have been attached in some previous parse state.

(289) a. That the food was good pleased me.
b. That John is tall is obvious.

Consider (289a). First, the word *that* is input. This word has at least two interpretations: a complementizer reading and a demonstrative determiner reading.[109] The load associated with the complementizer reading is $2x_{Int} + x_{TT}$ PLUs since 1) it is in a position that can receive a thematic role but does not as yet do so; 2) it has unsatisfied lexical requirements; and 3) it is a semantically null functional category in a thematic position and therefore will transmit a thematic role, but its lexical requirements contain no thematic elements yet. On the other hand, the load associated with the demonstrative determiner reading of *that* is 0 PLUs since this category has no lexical requirements and is not in a $\theta$-role receiving position.[110] Since the load difference between these two structures is so high, the complementizer reading of the word *that* is pruned and only demonstrative readings are pursued.

The word *the* is then input as a determiner. No attachments are possible and the parse cannot continue at this point. Backtracking results, and the word *that* is reanalyzed as a complementizer and then the parse continues successfully. While backtracking takes place in the analysis of sentences like those in (289), no attachments ever take place before backtracking is initiated. This reanalysis is therefore not expensive due to the Conscious Reanalysis Hypothesis. Thus no garden-path effect is predicted for the sentences in (289), as desired.

In contrast, when attachment of the demonstrative determiner reading of *that* is possible in a sentence initial position, a conscious garden-path effect is predicted. This prediction is confirmed by the data in (290) and (291) ((290d) from Milne (1986) and (291) from Fodor & Frazier (1980)):

---

[109] A third reading, the demonstrative pronoun reading, does not take part in the ambiguities considered here.

[110] Note that I am assuming that nouns are the heads of traditional noun phrases, in contrast with Abney's (1987) DP hypothesis.

(290) a. # That coffee tastes terrible surprised John.
b. # That gold is expensive is obvious.
c. # That furniture from IKEA will last a long time is likely.
d. # That deer ate everything in my garden surprised me.

(291) a. # That silly old-fashioned jokes are told too often is well-known.
b. # That badly-written poorly-acted movies still cost $6.00 to see annoys me.
c. # That tall people make better basketball players is clear.

Consider (290a). The initial portion of this sentence *that coffee tastes terrible* can be analyzed as either a simplex sentence, in which the word *that* is treated as a demonstrative determiner, or as a complementizer phrase that needs a thematic role. The first analysis is strongly preferred, and a garden-path effect results, since it is the CP reading that is necessary for a grammatical sentence.

The conscious difficulty that people experience with this sentence is as predicted by the Conscious Reanalysis Hypothesis inside the parallel parsing framework proposed here. First, the word *that* is input. Only demonstrative readings for this word are maintained, as described above. The noun *coffee* is then input and is attached as the head of a noun phrase. The input *tastes terrible* is then input and is also attached. When the verb *surprised* is input, no attachments are possible and the parse that has been pursued thus far fails, forcing reanalysis. This reanalysis is expensive since the input over which reanalysis takes place had been successfully analyzed up to that point: attachments have taken place over the initial input string.

The difficulty in processing the sentences in (291) is explained similarly. Consider (291a). The preferred representation for the input *that silly old-fashioned* is as a head-less noun phrase. When the plural noun *jokes* appears, no attachments are possible, and reanalysis is necessary. This reanalysis is expensive by the Conscious Reanalysis Hypothesis, since the input to be reanalyzed has representations that have already been attached. Hence the garden-path effects in (291) are also explained.

### 6.9.6. The Word *To*: Inflection Marker / Preposition Ambiguity

The word *to* is lexically ambiguous in English. It can either be a marker of inflection or a preposition. It turns out that in situations where both readings of *to* are permissible, both are retained, up to other constraints, as is illustrated by the examples in (292):

(292) a. I opened the letter to Mary.
b. I opened the letter to impress Mary.
c. I shouted to mow the lawn to Bob.
d. I shouted to Bob to mow the lawn.

Consider the parse of (292a) at the point of processing the word *to* (*cf.* Milne (1986)):

(293) a. [$_{IP}$ [$_{NP}$ I ] [$_{VP}$ opened [$_{NP}$ the letter [$_{PP}$ to ] ]]]
b. [$_{IP}$ [$_{NP}$ I ] [$_{VP}$ [$_{VP}$ opened [$_{NP}$ the letter ]] [$_{CP}$ [$_{IP}$ to ]] ]]

In structure (293a) the word *to* is attached to the NP *the letter* as a prepositional phrase. In structure (293b) the word *to* is attached to the VP *opened the letter* heading a purposive clause. The load associated with each of these structures is $x_{Int}$ PLUs, since in each structure the lexical requirements of the word *to* are unsatisfied. Thus there is no load difference between the two structures in each sentence, and both can be maintained. Hence the lack of garden-path effect for the sentences in (292) is predicted.

Now consider the sentences in (294), in which the verb *opened* in (292a) and (292b) is replaced with the verb *sent*:

(294) a. I sent the letter to Mary.
b. I sent the letter to impress Mary.

Unlike the verb *open* the verb *send* optionally subcategorizes for goal prepositional phrase. Hence in (294a) the preposition *to* can attach not only as modifier of the NP *the letter* but also as an argument to the verb *sent*. In most examples considered thus far, argument attachments are strongly preferred over adjunct attachments (see *e.g.*, (294a) in Section 6.3). Somewhat surprisingly, however, the attachment of *to* as a purposive modifier as in (294b) causes no conscious processing difficulty, despite the possibility of argument attachment. This effect is explained by the fact that although a structure representing the word *to* can attach as an argument in processing (294b), the structure that attaches as argument is of a different category than the purposive modifier clause and is therefore not comparable via the Property of Thematic Reception. Consider structures for the two relevant attachments:

(295) a. [$_{IP}$ [$_{NP}$ I ] [$_{VP}$ sent [$_{NP}$ the letter ] [$_{PP}$ to ] ]]
b. [$_{IP}$ [$_{NP}$ I ] [$_{VP}$ [$_{VP}$ sent [$_{NP}$ the letter ]] [$_{CP}$ [$_{IP}$ to ]] ]]

In structure (295a) the word *to* is attached to the VP headed by *sent* as a prepositional phrase argument. The load associated with this structure is $x_{Int}$ PLUs, since the preposition *to* has lexical requirements that need to be filled. Note that the lexical requirements of the verb *sent* are satisfied in this structure.

In structure (295b), as in structure (293b), the word *to* is attached as a purposive modifier of the VP. The load associated with structure (295b) is $2x_{Int}$ PLUs, since the lexical requirements of the preposition *to* and the verb *sent* are not yet satisfied. Note that there is no load via the Property of Thematic Reception associated with the purposive clause headed by *to* since this structure is not attached in a thematic position in any other coexisting representation.

Thus the load difference between the structures in (295) is $x_{Int}$ PLUs, not enough to cause pruning, and no garden-path effect is predicted in either of the sentences in (295), as desired.

### 6.10. The Properties of Thematic Reception, Lexical Requirement and Thematic Transmission: Solution Space

The following set of inequalities have been obtained thus far:[111]

(296) a. $x_{Int} \leq P$
b. $2x_{Int} > P$
c. $|2x_{Int} - x_{TT}| \leq P$
d. $|x_{Int} - x_{TT}| \leq P$
e. $x_{Int} + x_{TT} \leq P$

Inequalities (296c) and (296d) can each be expanded into two inequalities:

(297) a. $2x_{Int} - x_{TT} \leq P$
b. $x_{TT} - 2x_{Int} \leq P$
c. $x_{Int} - x_{TT} \leq P$
d. $x_{TT} - x_{Int} \leq P$

Furthermore, all values for $x_{Int}$ and $x_{TT}$ are assumed to be positive. It turns out that all of these inequalities are consistent. Hence it is possible to solve for the possible values associated with the relevant properties in terms of the preference constant $P$. The solution space of such values is depicted in Figure 11.

On this graph the loads associated with the Properties of Thematic Reception and Lexical Requirement are represented as a single variable, $x_{Int}$, on the horizontal axis, while the load associated with the Property of Thematic Transmission, $x_{TT}$, is represented on the vertical axis. Following simple linear programming techniques, the areas that satisfy all the relevant inequalities are then shaded on the graph, so that the intersection of all of these areas satisfies all of them. Thus every point in the shaded area of the graph in Figure 11 satisfies all the inequalities in (296). For example, the point $x_{Int} = 0.6P, x_{TT} = x_{Int}/2 = 0.3P$ lies inside the solution space and therefore satisfies all the inequalities.

Note that $x_{TT} < x_{Int}$ as a result of inequalities (296b) and (296e). Furthermore, note that many of the inequalities in (296) and (297) follow from the others. The inequalities that are crucial to the definition of the solution space are inequalities (296b), (296e) and (297a). Figure 12 gives a graph depicting the solution space for $x_{Int}$ and $x_{TT}$ in which only the crucial inequalities are represented.

[111]See Gibson (1990a) for a description of the results that are obtained when the loads associated with Properties of Thematic Reception and Lexical Requirement (Thematic Assignment in Gibson (1990a)) are assumed to be independent.

Figure 11: The Solution Space for the Inequalities in (296)

Figure 12: The Solution Space for Inequalities (296b), (296c) and (297a)

## 7. The Property of Recency Preference

Although the Properties of Thematic Reception and Lexical Requirement can account for a number of local preferences, they cannot account for the attachment preferences found in sentences like (298a) and (298b):

(298) a. Bill thought John died yesterday.
b. John figured that Sue wanted to take the cat out.

First consider (298a). The temporal adverb *yesterday* can attach in one of two possible sites: as modifier of the embedded verb verb *died* or as modifier of the matrix verb *thought*. The strongly preferred interpretation of this sentence links *yesterday* with the embedded verb. Neither the Property of Thematic Reception nor the Property of Lexical Requirement is relevant to the preference, since neither possible attachment deals with argument positions. The two possible structures are given in (299):

(299) a. $[_{IP}$ Bill $[_{VP}$ thought $[_{IP}$ John $[_{VP}$ $[_{VP}$ died $]$ $[_{AdvP}$ yesterday $]$ $]]]]$
b. $[_{IP}$ Bill $[_{VP}$ $[_{VP}$ thought $[_{IP}$ John $[_{VP}$ died $]]]$ $[_{AdvP}$ yesterday $]$ $]]$

Similarly, the strongly preferred reading of (298b) links the particle *out* to the embedded verb *take* rather than to the matrix clause *figured*. Although the particle *out* occupies an argument position, it does so in both structures, so that once again the Properties of Lexical Requirement and Thematic Reception must be irrelevant.

These effects were first noted in Kimball (1973). Kimball's explanation for the difficulty in obtaining reading (299b) for sentence (298a) was rooted in the *Principle of Right Association*. This principle stated that words are preferentially attached to the lowest nonterminal node in the structure built thus far for the sentence. Hence the adverb *yesterday* would be attached as a modifier of the more embedded verb phrase in (298a) because of the Principle of Right Association. Later work by Frazier & Fodor (1978) and Frazier (1979) gave a revised formulation of this principle known as one of *Local Association*, *Local Attachment* or *Late Closure* (Frazier (1979); Frazier & Fodor (1978); Frazier & Rayner (1982); see Section 2.2).

The intuition behind the principles that account for Late Closure effects is that incoming structures prefer to be associated with the words that occurred most recently in the sentence. The property that I present here, the Property of Recency Preference (PRP), makes this intuition explicit: one structure is associated with processing load via the PRP if there exists a structure resulting from a more recent attachment. That is, load is associated with maintaining links – thematic or non-thematic – over larger sections of the input string than is possible. Thus given the choice of matching a number of H-nodes, the PRP ensures that the most recent H-nodes are the preferred sites of attachment.[112]

[112] It might also be possible to formulate the PRP as a cost associated with a parser *operation* as opposed to a property of structure. That is, there might be a processing cost associated with making attachments to less recent words than is possible. There is then no load associated with the result of the attachment: it is the operation of attachment that is associated with cost. This alternative view of Recency Preference

(300) The Property of Recency Preference (PRP):
The load associated with the structure resulting from attachment involving a(n) (thematic, arbitrary) H-node $H$ on the stack = (number of more recent words that are associated with a matching (thematic, arbitrary) H-node) $*x_{RP}$ PLUs.

Note that the PRP is stated in terms of matching either *thematic* or *arbitrary* H-nodes. A thematic H-node is one that takes part in a thematic-role assigning relationship – thematic assignment, thematic transmission or thematic reception – when it is realized. An *arbitrary* H-node is any H-node at all, including thematic H-nodes. Thus if the most recent possible attachment site is a thematic H-node, then all less recent attachment sites are then associated with load via the Property of Recency Preference. If, however, the most recent possible attachment site is not a thematic H-node, then this site affects only those less recent sites that are also non-thematic sites. Less recent thematic attachment sites will not be affected.

The distinction between thematic H-nodes and other H-nodes is also made in the statement of the Property of Thematic Reception, (218). Theoretical motivation for this distinction comes from the $\theta$-Criterion: the parser wants to locally satisfy the $\theta$-Criterion whenever possible. That is, assignment of thematic roles takes top priority when choosing between multiple attachment sites. Empirical consequences of this will be given below.

In order to see how the Property of Recency Preference functions, consider once again (298a) together with the structure for this sentence when the word *yesterday* is input:

(298a) Bill thought John died yesterday.

(301) [$_{IP}$ Bill [$_{VP}$ [$_{VP}$ thought [$_{IP}$ John [$_{VP}$ [$_{VP}$ died ] [$_{AdvP}$ *H-node*$_{died}$ ] ]]] [$_{AdvP}$ *H-node*$_{thought}$ ] ]]

There are two possible attachment sites for the adverb *yesterday*: either as a modifier of the lower clause, or as a modifier of the matrix clause. These sites are indicated by H-node AdvP nodes adjoined to each VP, labeled *H-node*$_{died}$ and *H-node*$_{thought}$ respectively.[113] The load associated with attaching *yesterday* as an adjunct to the lower VP (matching *H-node*$_{died}$) in (301) is $0 * x_{RP}$ PLUs, since the H-node being matched in this attachment is associated with the most recent word in the sentence. The load associated with the structure resulting from attaching *yesterday* as an adjunct to the matrix verb *thought* (matching *H-node*$_{thought}$) is $= 1 * x_{RP}$ PLUs, since there is one more recent word that is associated with an matching H-node.

---

implies a race-based implementation of the framework presented here (Frazier & Fodor (1978), Fodor & Frazier (1980), McRoy & Hirst (1990)). That is, given the empirical effects to be accounted for in this section (see *e.g.*, (337) and (335)), it must be that structures resulting from less recent attachments are only ever considered if the load associated with their results can be less than the load associated with the result of structures resulting from more recent attachments. Such an implementation seems possible, and would make the same empirical predictions as the model presented here.

[113] It is possible that the adjunct sites are at the IP level. However, it makes no difference which of these hypotheses is followed for the arguments to be presented here. Thus, without loss of generality, I will assume that adverbs of time are adjoined at the VP level (*cf.* Ford, Bresnan & Kaplan (1982)).

Thus the load associated with matching *H-node*$_{died}$ is $x_{RP}$ PLUs, whereas the load associated with matching *H-node*$_{thought}$ is 0 PLUs. Since there are no other differences between the two attachments, the load difference between the two resulting structures is $x_{RP}$ PLUs. Following the intuitions of other psycholinguists (see, for example, Kimball (1973), Frazier (1979) and Frazier & Fodor (1978)) I assume that this preference for associating *yesterday* with the most deeply embedded clause in (298a) is a strong one; that is, the alternative representation that links *yesterday* to the matrix clause is not pursued. Thus I assume that the load difference between structures (299a) and (299b) is large enough to cause pruning:

(302) $x_{RP} > P$

The existence of the preferred reading in (298b) is explained in a similar way:

(298b) John figured that Sue took the cat out.

Consider the two possible attachment sites of the particle *out*:

(303) [$_{IP}$ [$_{NP}$ John ] [$_{VP}$ [$_{V'}$ [$_{V}$ figured ] [$_{CP}$ that [$_{IP}$ [$_{NP}$ Sue ] [$_{VP}$ [$_{V'}$ [$_{V}$ took ] [$_{NP}$ the cat ] [$_{ParticleP}$ *H-node*$_{took}$ ] ]]]] [$_{ParticleP}$ *H-node*$_{figured}$ ] ]]]

The load associated with the structure that results from attachment to *H-node*$_{took}$ is O PLUs since there are no attachment sites associated with more recent words. Furthermore, the load associated with matching *H-node*$_{figured}$ is $x_{RP}$ PLUs, since there exists an attachment site associated with a more recent word in the input. Thus the load difference between the structures resulting from these attachments is $x_{RP}$ PLUs, and a preferred reading is predicted, as desired.

### 7.1. Node Closure

While the Property of Recency Preference makes the right predictions with respect to the preferred readings in sentences like those in (298), by itself it makes an incorrect prediction with respect to the processing of right branching structures like those in (304) ((304a) from Church (1980)):

(304) a. This is the dog that chased the cat that ran after the rat that ate the cheese that Mary bought at the store...

b. I saw the mouse on the box in the old shed by the house on the lake near the highway...

The first relative clause in (304a) has only one possible attachment site: as modifier to the NP *the dog*. The second relative clause has two possible attachment sites: as modifier to either the NP *the cat* or the NP *the dog*. Similarly, the third relative clause has three possible attachment sites, and so on. Given the Property of Recency Preference, each of these possible attachment sites must be evaluated with respect to each attachment. Since short term memory is limited, there must be a limit on the number of possible attachment sites that can be considered in parallel by the human parser. Thus the Property of Recency Preference by itself predicts that at some point right branching structures should get markedly difficult to comprehend, just like center-embedded structures. This is clearly not the case, as is demonstrated by the lack of difficulty with the sentences in (304).

This problem with the Property of Recency Preference is similar to a difficulty with Frazier's (1979) principle of Late Closure observed by Church (1980). The Principle of Late Closure tells the parser how to select between attachment sites. However, it does not tell the parser how to remove attachment sites from consideration, thereby closing phrases and easing the short term memory cost. Kimball (1973) proposed the principle of Closure as one way to limit future attachment choices:

(305) Closure: A phrase is closed as soon as possible, *i.e.*, unless the next node parsed is an immediate constituent of that phrase. [Kimball (1973), p. 36]

Kimball's principle of Right Association predicts that the the preferred reading of (298a) links the adverb *yesterday* to the lower clause. The principle of Closure dictates that once this attachment has been made, the higher clause is no longer a possible attachment site: it has been closed off. While Kimball's Closure principle solves the memory problem noted by Church, it has many empirical problems. For example, the sentences in (306) and (307) are predicted to be difficult to understand, but none of them poses any problem to the human processor ((306a) is from Church (1980); the sentences in (307) are from Frazier & Fodor (1978)):

(306) a. I called the guy who smashed my brand new car a rotten driver.
b. I called the guy in the parking lot a rotten driver.
c. The policeman persuaded the man who admitted his guilt to turn himself in.

(307) a. We went to the lake to swim but the weather was too cold.
b. Grandfather took a long hot bath after he had finished his chores.

Consider (306a) and (306b). Once a modifier is attached to the NP *the guy* (the relative clause *who smashed my brand new car* in (306a); the PP *in the parking lot* in (306b)), the verb phrase headed by *called* closes according to the principle of Closure, (305). Thus, (306a) and (306b) should give the human processor difficulty, but no difficulty is apparent in either. Similarly, (306c) is incorrectly predicted to be problematic.

Kimball's Closure principle also predicts that the matrix sentence level node is closed in (307a) when the sentential adjunct *but the weather was too cold* arrives as input. As a result, this sentence should also be difficult to process, but it is not. Similarly, (307b) is incorrectly predicted to be difficult by (305).

As a result of the theoretical difficulty with Frazier's principle as well as the empirical difficulty with Kimball's principle, Church proposed a compromise: the *A-over-A Early Closure Principle*:

(308) The A-over-A Early Closure Principle:
Given two phrases in the same category (*e.g.*, noun phrase, verb phrase, clause, *etc.*), the higher closes when both are eligible for Kimball closure. That is, 1) both nodes are in the same category, 2) the next node parsed is not an immediate constituent of either, and 3) the mother and all obligatory daughters have been attached to both nodes. [Church (1980), p. 33]

Church's principle is like Kimball's Closure principle in that it closes nodes for attachment under certain conditions. But the conditions for closure are much more restrictive, so that it avoids many of the empirical difficulties associated with Kimball's principle. Consider (306a) with respect to the A-over-A Early Closure Principle. The verb phrase headed by *called* never closes under this closure principle despite the fact that a second verb phrase appears, because the second VP never becomes eligible for Kimball Closure. Similarly, none of the other sentences in (306) or the sentences in (307) is predicted to be difficult under Church's closure principle, as desired. Furthermore, Church's principle closes many of the early attachment sites found in the right branching structures in (304) so that the memory limitation problem associated with Frazier's Late Closure principle is also avoided.

In order to circumvent the memory limitation difficulty in the framework presented in this thesis, I propose to import a principle of closure similar to Church's. While Church's closure principle is a great improvement over Kimball's, it still suffers some empirical problems. These problems stem from the fact that a node whose obligatory arguments are satisfied can become closed as long as a second node is present whose syntactically obligatory arguments are also satisfied, without regard for the type of syntactically optional arguments each node can take. Consider the sentences in (309) and (310):

(309) a. I convinced the woman who John loves very much that she should call him.
b. The teacher persuaded the student who had an extremely serious drug problem to seek counseling.
c. The superintendent talked to the boy that owns three dogs about the latest incident.
d. The waitress talked to the customer that ate the pie about the coffee.
e. The teller transferred the money that I had been saving to my new bank.
f. I convinced the man who dumped the trash to clean it up.

(310) a. I convinced the woman who I think John loves very much that she should call him.
b. The teacher persuaded the student who admitted that he had a drug problem to seek counseling.
c. The superintendent talked to the boy that owns the dog that is always chasing cats about the latest incident.
d. The waitress talked to the customer that ate the piece of pie about the coffee.
e. The teller transferred the money that I had been keeping in my old bank to my new bank.
f. I convinced the man who dumped the trash on the floor to clean it up.

Church's closure principle predicts that each of the sentences in (310) is markedly more difficult than its counterpart in (309). This prediction seems incorrect to me and my informants. Consider (310a). This sentence is predicted difficult by Church's A-over-A Early Closure Principle because at the point that the adverbial modifier *very much* is being processed the verb phrases headed by *convinced* and *think* both satisfy Kimball closure. That is, both of these verbs have all of their obligatory arguments satisfied, and the input currently being processed is not an immediate daughter of either. Thus the higher VP, headed by *convinced*, is closed off, and processing difficulty should immediately ensue when the CP *that she should call him* follows. Similarly, (310e) is predicted difficult, since both of the VPs headed by *transferred* and *keeping* satisfy Kimball closure when the NP *my old bank* is being attached as object of the preposition *in*. Thus processing difficulty should result when the PP *to my new bank* is input, since this PP must attach to a closed VP. However, people have no difficulty with these kinds of sentences, so Church's principle is too strong.

In order to avoid the empirical problems associated with Church's principle, I propose a principle of closure that removes attachment sites from consideration only when a comparable attachment site is available that would be significantly less costly. Two attachment sites are said to be *closure-comparable* if 1) they are of the same $\overline{\mathrm{X}}$ category and 2) their maximal projections are attached to the same kind of $\overline{\mathrm{X}}$ category. The Principle of Preference Closure is defined in terms of closure-comparability:

(311) The Principle of Preference Closure (PPC):
Given a structure, prune off all optional attachment sites in that structure whose processing loads after attachment would differ from a closure-comparable optional attachment site in that structure by at least the Preference Factor $P$ PLUs.

In other words, consider all possible optional attachment sites of category $C_1$ (H-nodes) dependent upon category $C_2$ in some structure $S$. Pick one of these, $A$, that would be associated with a minimal processing load via all properties that currently apply.[114] Prune all attachment sites in $S$ that would be associated with a processing load that is at least $P$ PLUs more expensive than the cheapest possible, $A$.

The Principle of Preference Closure works in a manner very similar to Church's A-over-A Early Closure Principle. Consider (298a) once again:

(298a) Bill thought John died yesterday.

At the point of processing the verb *died* in (298a) the structure in (301) has been built:

(301) [$_{IP}$ Bill [$_{VP}$ [$_{VP}$ thought [$_{IP}$ John [$_{VP}$ [$_{VP}$ died ] [$_{AdvP}$ *H-node$_{died}$* ] ]]] [$_{AdvP}$ *H-node$_{thought}$* ] ]]

There are two H-node AdvP attachment sites in this structure, indicated as *H-node$_{died}$* and *H-node$_{thought}$*. The load associated with attachment to *H-node$_{died}$* is 0 PLUs, while the load associated with attachment to *H-node$_{thought}$* is $x_{RP}$ PLUs. Since these two attachment

[114]Note that lexical, semantic, pragmatic and contextual properties will not normally differentiate attachment sites at this point since the relevant properties of the item to be attached are not yet known.

sites are both dependent on the same type of category, a verb phrase, they can be compared by the Preference Closure Principle. The attachment site $H\text{-}node_{thought}$ is then pruned from the structure, since its potential processing load differs by more than the Preference Factor $P$ PLUs from that of another attachment site. Similarly, early attachment sites are pruned from consideration in the right branching structures in (304) when later similar attachment sites appear. Thus human short term memory limitation no longer blocks right branching structures from being processed.

Furthermore, the PPC does not suffer from the empirical disadvantages of Kimball's or Church's Closure principles that were noted earlier. Once again consider (306a):

(306a) I called the guy who smashed my brand new car a rotten driver.

When the verb *smashed* is input in (306a), adjunct attachment sites are pruned from the higher level VP *called*, but the NP argument H-node associated with *called* is not pruned, since the verb *smashed* does not take an optional NP argument. Thus this attachment site is still available when the NP *a rotten driver* appears, and attachment takes place, as desired. Similarly, the PPC correctly predicts no difficulty with sentences like (310a):

(310a) I convinced the woman who I think John loves very much that she should call him.

When the verb *think* is input, adjunct VP attachment sites associated with the verb *convinced* are pruned by the PPC, but the optional argument CP attachment site remains since the verb *think* does not take an optional CP.[115] Thus this site is available when the CP *that she should call him* arrives, and no processing difficulty is experienced. Similarly, the other sentences in (310) do not give the Principle of Preference Closure any difficulty, as desired.

The Principle of Preference Closure has many further empirical consequences. In particular, since lexical, semantic, pragmatic and contextual properties cannot distinguish between attachment sites without some information about the item to be attached, the PPC predicts that some attachments will be performed independent of these influences. This seems to be the correct prediction, as is evident in the examples in (312):

(312) a. # Bill thought John would die yesterday.
b. # John figured that Sue took the train to New York out.

Despite the facts that 1) the adverbial *yesterday* semantically must modify a past tense verb and 2) there exists a past tense verb phrase attachment site, *yesterday* initially attaches to the future tense verb phrase *will die* in (312a), giving an initial semantically bizarre interpretation. Because of the semantic anomaly of the resulting structure, reanalysis takes place and the adverbial re-attaches as modifier of the past tense verb *thought*. This garden-path effect is as predicted by the combination of the Property of Recency Preference and the Principle of Preference Closure. When the verb *die* is input, the Principle of Preference Closure applies, pruning H-node adjunct attachment sites from the VP *thought* because of

[115] One reading of the verb *think* takes an obligatory CP argument, but none takes an optional one.

processing load evaluation differences, which result from the Property of Recency Preference. Since the VP attachment sites are removed at this point, the only available attachment site is as modifier of the VP *die*. When the adverbial *yesterday* is then input, it must attach to this position, resulting in the strange initial interpretation of (312a). Similarly, (312b) is initially interpreted with the bizarre interpretation which results when the particle *out* modifies the lower clause.

So far, the examples that have been considered have mainly involved ambiguities in verb phrase attachment. The Property of Recency Preference and Principle of Preference Closure also make predictions with respect to NP attachment sites. In particular, The PRP and PPC combine to predict that the sentences in (314) are significantly more difficult to process than their counterparts in (313):[116]

(313) a. The pots beside the plant next to the tables that are wooden fell to the floor.
b. I liked the woman with the dog by the apartment buildings that are made of stone.
c. George drank the water which came from the jar on the napkins on the table that is by the door.
d. The tree by the house on the lake named Windermere is one hundred years old this year.
e. The apple near the cup on the table painted red had a bite taken out of it.

(314) a. # The pots beside the plant next to the tables that is flowering fell to the floor.
b. # I liked the woman with the dog by the apartment buildings that was barking all night long.
c. # George drank the water which came from the jar by the napkins on the table that are white with red stripes.
d. # The tree by the house on the lake built from brick is one hundred years old this year.
e. # The apple near the cup on the table removed from the dishwasher had a bite taken out of it.

Consider (313). The CP *that was flowering* must attach as a modifier of the NP *the plant* because of subject verb number agreement, but people have conscious difficulty in making this attachment. This is as predicted by the Property of Recency Preference and the Principle of Preference Closure. When the NP *tables* is input, the NP *plant* is closed off by the PPC, since all of its associated attachment sites have more recent counterparts linked with *tables*. When the CP *that was flowering* is input, it cannot attach to the NP *tables*

[116]The NP recency preference garden-path effects in (314) are not as strong as those associated with many other garden-path sentences. However, people still experience conscious difficulty with the examples in (314). The theory given here predicts only the fact that reanalysis, corresponding to conscious difficulty, takes place. It does not predict how the reanalysis occurs, which would explain the difference in degree of effect among these examples. Presumably, the NP recency preference examples are easier to reanalyze than many other garden-path examples, since all the same structure can be used in the reanalysis. However, no formalization of such a theory is presented here. See Frazier & Rayner (1982) for a discussion of some of the issues involved in reanalysis.

because of number disagreement, and a garden-path effect results. No such similar effect occurs in (313a) since all modifier attachments are to the most recently occurring head. A similar explanation exists for the distinction between each of the relevant examples in (313) and (314).[117]

While the PPC makes many predictions when the attachment site and head category are the same, it does not apply in examples like those in (317):

---

[117]None of the crucial attachment ambiguities in the sentences in (314) involves an argument of a verb. Verbal arguments in English, especially subjects, play important discourse roles that may interact with attachment preferences. That is, while structural properties play an important role in determining linguistic preferences, discourse properties, which may favor verbal argument attachment sites, especially subject positions, over other sites. This discourse property may then interact with other properties to give effects different from those predicted by structural properties like the PRP alone. For example, consider (315):

(315) The book on the table with 300 pages is sitting next to the teapot.

If only the Property of Recency Preference is considered, (315) is incorrectly predicted to induce a garden-path effect. That is, once the noun phrase *the table* is input, the NP *the book* should be closed for further attachments. When the PP *with 300 pages* is then input, processing difficulty is predicted since the only available attachment site is the NP *the table*: reanalysis is necessary to link this PP to a more pragmatically appropriate site, *the book*. However, there is no conscious difficulty in the processing of (315). Thus either the PRP is incorrect as it stands or there are other unconsidered properties interacting with the PRP in (315). I hypothesize the latter. That is, I hypothesize that there is a discourse-related property which prefers the matrix subject over the PP object as an attachment site in (315), so that the matrix subject site is not initially pruned. I propose (316) as an initial approximation to this property:

(316) Verbal Argument Attachment Preference Property: Associate a load of $x_{verb-pref}$ PLUs with a structure that results from the non-argument attachment of one structure to another, where the attachment site is not a verbal argument.

Consider the attachment of the preposition *with* in the processing of (315). The attachment of this preposition to the NP headed by *book* is associated with load via the Property of Recency Preference, but is not associated with any load via the Verbal Argument Attachment Preference Property. In contrast, the attachment of the preposition *with* to the NP headed by *table* is associated with no load via the Property of Recency Preference, but is associated with load via the Verbal Argument Attachment Preference Property. Thus the Verbal Argument Attachment Preference Property counteracts the Property of Recency Preference in certain circumstances, so that attachment to either the NP headed by *book* or the NP headed by *table* is possible in the processing of the preposition *with* in (315).

As a result of the interaction of a discourse related property like the Verbal Argument Attachment Preference Property, I have given examples that do not involve variation in verbal arguments attachment site status. Exactly what discourse-related processing properties and loads are appropriate is a topic for future research.

Furthermore, note that the load associated with a discourse related property like the Verbal Argument Attachment Preference Property may vary across languages. For example, Cuetos & Mitchell (1988) discuss evidence from Spanish which indicates that people prefer to attach following modifiers to the first noun phrase (the head) in possessive constructions (see also Mitchell, Cuetos & Zagar (1990)). They conclude that Recency Preference (Late Closure) is not active in all constructions of all languages. However, from their results it would be just as reasonable to conclude that Recency Preference remains constant (as they assume it does in other Spanish constructions) while the load(s) associated with discourse properties vary to give the appropriate effect for possessive constructions. Further experimentation is of course needed to see if this is a viable explanation.

(317) a. I saw the man in the park.
b. John carried a package for Susan.
c. Bill ate the food on the table.

The PPC applies only when the attachment sites are dependent on nodes of the same category. Thus in (317a), the PPC does not close the PP attachment site on the VP headed by *saw* when the NP *the man* is input, despite the presence of a more recent attachment site. The categories of the head nodes associated with the two hypothesized PPs differ, so that no attachment site closure occurs. The parsing preferences in the sentences in (317) are then determined by a combination of the PRP together with lexical, syntactic, semantic and pragmatic properties. See Ford, Bresnan & Kaplan (1982), Schubert (1984, 1986), Wilks, Huang & Fass (1985), Kurtzman (1985), Taraban & McClelland (1988, 1990), Stowe (1989) for further examples in which non-syntactic properties have significant effects. Also see Sections 9.2 and 9.3 for a description of some of this data as well as a partial account of it in the framework proposed here.

### 7.2. Thematic vs. Non-thematic Attachments

The Property of Recency Preference is stated in terms of matching either *thematic* or *arbitrary* H-nodes. There are a number of empirical consequences of this distinction. First, consider sentence (318):

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

As was noted in the Section 6.3, this sentence induces a garden-path effect. The local ambiguity in this sentence begins with the preposition *on*. This preposition may attach as an argument of the verb *put* or as a modifier of the NP *the candy*:

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

If we consider only the Properties of Thematic Reception and Lexical Requirement, the load associated with structure (319a) (argument attachment) is $x_{Int}$ PLUs and the load associated with structure (319b) (adjunct attachment) is $3x_{Int}$ PLUs.[118] Now let us consider the processing loads of each of these structures with respect to the Property of Recency Preference. First consider (319b). The preposition *on* modifies the NP *the candy*, the most recent constituent in the input string. Thus (319b) is not associated with any load by the PRP. Consider now structure (319a). The preposition *on* matches an argument requirement of the verb *put*. Since this is a thematic position, only thematic positions need be considered in calculating the load due to the Property of Recency Preference. Since there are no other thematic positions that can be matched by the preposition *on* at this stage in the parse, there is no load associated with the PRP for structure (319a) either. Thus the load calculations for (318) remain unchanged.

[118] See Chapter 5 for explanations of these loads.

Now consider the sentences in (320):

(320) a. # I put the candy that John saw on the table into my mouth.
b. # I put the book, the pencil and the eraser on the floor onto the table.

Like (318), the examples in (320) induce garden-path effects. Consider (320a). As in (318), the preposition *on* can either attach as argument of the verb *put* or as modifier of the verb *saw*. The argument attachment is strongly preferred, and hence a garden-path effect results. This effect is as predicted by the Properties of Thematic Reception, Lexical Requirement and Recency Preference.[119] Compare the examples in (320) to those in (322):[120]

(322) a. I put the candy that John threw on the table into my mouth.
b. I gave the furniture that Karen sent to the dormitory to the Salvation Army.

Unlike (320), (322) is not a garden-path sentence. The difference between (322) and (320) is that the verb *threw* optionally subcategorizes for a locative prepositional phrase, while the verb *saw* does not. Consider two possible structures of the input string *I put the candy that John threw on*:

(323) a. I $[_{VP}$ $[_{V'}$ $[_{V}$ put $]$ $[_{NP}$ the candy that John threw $]$ $[_{PP}$ on $]$ $]]$
b. I $[_{VP}$ $[_{V'}$ $[_{V}$ put $]$ the $[_{N'}$ $[_{N'}$ candy $]$ $[_{CP}$ that John $[_{VP}$ $[_{V'}$ $[_{V}$ threw $]$ $[_{PP}$ on $]$ $]]]]]]$

It turns out that loads associated with the Properties of Thematic Reception and Lexical Requirement are the same for each of the structures in (323), since both attachments are argument attachments. However, the Property of Recency Preference distinguishes the two structures. Matching the H-node argument PP of the verb *threw* does not force any load by the PRP since *threw* is the most recent word in the input string. On the other hand, matching the H-node argument PP of *put* forces a load of $x_{RP}$ on this structure, since there is a more recent thematic H-node to match. Thus structure (323b) is preferred and no garden-path effect results.

Similarly, the PRP predicts garden-path effects in the examples in (324) and predicts no garden-path effects in the examples in (325) ((324a) and (325a) from Pritchett (1988)):[121]

---

[119] Contrary to the predictions of the theory presented here, the framework of Frazier & Fodor (1978) predicts that the final PP in a sentence like (321) (from Milsark (1983)) will be attached to the immediately preceding noun rather than to the matrix verb:

(321) Susan bought a house with two bedrooms, a garage, and a fireplace for Mary.

However, Milsark (1983) disputes their judgment: he, along with his informants, prefer the reading of (321) that links the PP *for Mary* to the verb *bought*. All of the informants that I have questioned agree with Milsark on this issue. Note that the theory presented here predicts Milsark's judgments, since the attachment of *for Mary* to *bought* is a thematic attachment, while the attachment to the NP *fireplace* is not.

[120] Note that the Principle of Preference Closure does not apply in these sentences since the second argument of each of the matrix verbs is obligatory.

[121] Pritchett (1988) claims that the sentences in (324) and (325) cause similar effects in speakers: seemingly randomly inducing a garden-path effect or not. My informants do not agree with his judgment: they consistently experienced garden-path effects in sentences like those in (324), and consistently did not experience

(324) a. # Sue gave the man who was reading the book.
b. # John sent the girl who was eating the candy.

(325) a. Sue gave the man who was reading the book the magazine.
b. John sent the girl who was eating the candy a present.

Consider (324a). When the noun phrase *the book* is encountered, it may attach either as argument of the verb *reading* or as argument of the verb *gave*. Since both are $\theta$-attachments, the Properties of Thematic Reception and Lexical Requirement are irrelevant. However, the Property of Recency Preference distinguishes the tow possible attachments, preferring the attachment to the verb *reading*. Since it is the structure that results from the less recent attachment that is necessary in order to obtain a grammatical parse of (324a), a garden-path effect results. Similarly, the sentences in (325) are not difficult to process.

A further empirical consequence of distinguishing thematic and non-thematic H-nodes in the statement of the PRP is presented by examples like those in (326):

(326) a. # The horse that John saw beside the dog raced past the barn fell.
b. # The canoe that Susan observed beside the sailboat floated down the river sank.
c. # The dog that was fed next to the cat walked to the park chewed the bone.

The sentences in (326) induce strong garden-path effects. Consider (326a). The relevant local ambiguity begins at the word *raced*: structures for this word may either attach as the matrix verb of the sentence, or as modifier of the NP *the dog*. The modifier attachment is the one that is necessary to avoid a garden-path effect, but it turns out that the matrix verb attachment is strongly preferred, despite the fact that this attachment involves matching an H-node that is much older (less recent) than other available H-nodes. Note that the attachment of *raced* as the matrix verb of the sentence is a thematic attachment, while the attachment of *raced* as adjunct of the NP *the dog* is not a thematic attachment. Thus this garden-path effect is empirical evidence in support of the distinction between thematic and non-thematic H-node matches in the statement of the PRP.

## 7.3. Recency Preference: Interactions with the Properties of Thematic Reception and Lexical Requirement

Four properties have been proposed thus far: the Property of Thematic Reception, the Property of Lexical Requirement, the Property of Thematic Transmission and the Property of Recency Preference. It has been shown in the previous two sections how these properties act independently to predict garden-path effects. An obvious extension of this work is to see how these properties interact. This section will illustrate some of the interactions among all

---

garden-path effects in sentences like those in (325).

of the properties presented here.[122]

### 7.3.1. Direct Object / Matrix Clause Subject Ambiguities

Consider the sentences in (327) ((327a) from Frazier & Rayner (1982)):

(327) a. # Since she jogs a mile seems light work.
b. # While the cannibals ate the missionaries drank.
c. # Since Jane married John dated others.

These sentences are locally ambiguous at the point of parsing the post verbal noun phrase. Since each of the verbs in the initial clause is ambiguous between transitive and intransitive readings, the post-verbal NP may attach as either the object of the verb or as subject of the matrix clause to come. The object attachment is strongly preferred locally and a garden-path effect results in each of the examples in (327) since it is the matrix subject attachment that is necessary for a grammatical parse.[123] The strong preference for the direct object attachment will be explained here by appeal to the Properties of Recency Preference and Thematic Reception. Consider (327a) at the point of parsing the NP *a mile*. Attaching *a mile* as object of the verb *jogs* involves matching an H-node associated with the most recent word, while the attachment as matrix subject involves matching H-node associated with a less recent word. Furthermore, the resulting matrix subject does not receive a thematic role, while the same NP receives a thematic role as object of *jogs*. Thus the direct object reading is strongly preferred. A formal derivation of the processing load difference is given below. Consider two structures for the input *since she jogs*:

(328) a. $[_{IP}$ $[_{CP}$ since $[_{IP}$ $[_{NP}$ she $]$ $[_{VP}$ $[_{V'}$ $[_{V}$ jogs $]$ $[_{NP}$ $H\text{-}node_1$ $]$ $]]]]$ $[_{IP}$ $H\text{-}node_2$ $]]$
b. $[_{IP}$ $[_{CP}$ since $[_{IP}$ $[_{NP}$ she $]$ $[_{VP}$ jogs $]]]$ $[_{IP}$ $H\text{-}node_3$ $]]$

Structure (328a) represents the transitive reading of *jogs* while (328b) represents the intransitive reading. Thus the verb *jogs* in (328a) is followed by an H-node NP, labeled $H\text{-}node_1$. Note that both representations include matrix IP H-nodes, labeled $H\text{-}node_2$ and $H\text{-}node_3$ respectively. These H-nodes indicate that a matrix clause must follow both representations.

Consider now representations for the NP *a mile*:[124]

---

[122]Since thematic transmission only applies to semantically null categories and semantically null categories are less common than those with some semantic contribution, the Property of Thematic Transmission does not apply as often as the other properties discussed here. As a result, I have not yet found evidence in which the Property of Thematic Transmission interacts with the Property of Recency Preference. I leave investigation of these interactions, to future research.

[123]There is some question in the literature as to whether or not sentences like those in (327) and (333) (to come) are grammatical. See footnote 20 for arguments that the difficulty experienced in processing these sentences is not due to the grammar and therefore must be due to the processor.

[124]The parser will of course encounter the determiner *a* first, and then encounter the word *mile*. However,

(329) a. $[_{NP}$ a mile $]$

b. $[_{XP_{clause}}$ $[_{NP}$ a mile $]$ $]$

Structure (329a) is the bare NP representation for *a mile*. Structure (329b) includes an H-node *clausal* phrase, indicating that an NP can be the subject of a clause.

Attachments are now made between the structures in (328) and the structures in (329). Two attachments are possible: the H-node NP object of *jogs*, *H-node*$_1$, in (328a) can be satisfied by attaching structure (329a); and the H-node matrix IP, *H-node*$_3$, in (328b) can be matched with the clausal H-node in (329b). These two attachments result in the following two structures:

(330) a. $[_{IP}$ $[_{CP}$ since $[_{IP}$ $[_{NP}$ she $]$ $[_{VP}$ $[_{V'}$ $[_{V}$ jogs $]$ $[_{NP}$ a mile $]$ $]]]]$ $[_{IP}$ $]]$

b. $[_{IP}$ $[_{CP}$ since $[_{IP}$ $[_{NP}$ she $]$ $[_{VP}$ jogs $]]]$ $[_{IP}$ $[_{NP}$ a mile $]$ $]]$

Consider the processing loads of these two structures with respect to the Property of Recency Preference. Structure (330a) is associated with no load by the PRP, since the H-node involved in the last attachment, *H-node*$_1$, was the most recently built H-node matched. Structure (330b), on the other hand, is associated with $x_{RP}$ PLUs, since the attachment to produce this structure involves *H-node*$_3$, an H-node that was built before *H-node*$_1$.

Now consider the processing loads of these two structures with respect to the Properties of Thematic Reception and Lexical Requirement. Structure (330a) is associated with no processing load by either the PTA or the PTR. Structure (330b) is associated with $x_{Int}$ PLUs since the NP *a mile* is in a $\theta$-position, but does not yet receive a thematic role. Note that the verb *jogs* has an intransitive reading, so that all thematic roles are assigned inside this representation. Thus the total load associated with (330a) is 0 PLUs, while the load associated with (330b) is $x_{RP} + x_{Int}$ PLUs. Since (327a) is a garden-path sentence, we have:

(331) $x_{Int} + x_{RP} > P$

This inequality is in fact redundant since it was already established that $x_{RP} > P$ in (302). Thus the garden-path effect in (327a) is predicted by the Properties of Thematic Reception, Lexical Requirement and Recency Preference. [125]

---

nothing of relevance to the presentation at hand happens when the determiner *a* is input. Thus I continue the presentation at the word *mile*. The results as presented here remain unchanged when all parse states are considered.

[125] An alternative explanation of the garden-path effects in (327) says that these effects are due to a parsing strategy that prefers attachment sites that cannot contain a comma in points of ambiguous attachment where a comma is possible (Kurtzman 1985). This theory also assumes that the placement of commas does not rely on local attachment preferences, but is determined independently. (If this theory assumed that the placement of commas depended on local attachment preferences, then it would reduce to a theory of local attachment preferences.)

A theory which predicts that comma placement is independent of local preferences predicts that given two attachment sites, one which can include a comma and one which cannot, no matter how lowly ranked the attachment site that cannot contain a comma is, with respect to local attachment preference metrics, it will always be the preferred attachment site. For example, the following sentences should therefore induce

### 7.3.2. Lexical Ambiguity and Recency Preference: Subject / Object Ambiguity

Consider the sentences in (333) ((333a) from Frazier (1979)):

(333)a. # Without her contributions would be inadequate.
b. # Since developing the computer companies in the United States have grown enormously.
c. # Having seen the Martian people in the town panicked.

The sentences in (333) invoke garden-path effects, just as those in (327) do. The explanation is very similar. Consider (333a). The local ambiguity in (333a) begins with the word *her* and ends at the word *contributions*:

(334)a. $[_{IP}$ $[_{PP}$ without $[_{NP}$ $[_{NP}$ her $]$ $[_{N'}$ contributions $]]]$ $[_{IP}$ $]]$
b. $[_{IP}$ $[_{PP}$ without $[_{NP}$ her $]]$ $[_{IP}$ $[_{NP}$ contributions $]$ $]]$

Structure (334a) has the genitive reading of *her* attached as specifier of the NP headed by *contributions*. This NP is attached as direct object of the preposition *without*. Structure (334b) has the accusative reading of *her* attached as object of the preposition *without*. The NP *contributions* is the subject of the matrix clause to follow. Reading (334a), is strongly preferred, and a garden-path effect results for (333a) since reading (334b) is the one that leads to a successful parse.

The explanation of the garden-path effect for (333a) is similar to that for (327a). When the NP *contributions* is read, it can either attach as direct object of *without*, or as subject of the matrix clause. Consider these possible attachments with respect to the Property of Recency Preference. There is no PRP load associated with the direct object attachment, since this attachment involves matching an H-node associated with the most recent word in the input string, *her*. The attachment as matrix subject, on the other hand, is associated with $x_{RP}$ PLUs since this attachment involves matching an H-node that is associated with the word *without*, a less recent word. Furthermore, structure (334b) contains an NP *contributions*

---

garden-path effects just as strong as (27a):

(332)a. After I bought the dog Mary became very jealous.
b. Since Fred brought the salad Wilma felt obliged to eat it.
c. The last time I went to the store soap was very expensive.
d. Whenever he smokes his pipe women avoid him.
e. While I was sleeping the cat crept onto my bed.

For example, consider (332a). The verb *buy* can take two objects so that the input string *After I bought the dog Mary* is syntactically ambiguous: the NP *Mary* may either be the second object of the verb *bought* or it may be the matrix subject of the verb to follow. There is a strong pragmatic preference for the matrix subject interpretation of the NP *Mary*, since it is very unlikely that a person would buy another person for a dog. Even so, both readings are possible. The theory of preferences that says that structures in which commas are impossible are always preferred predicts that (332a) should induce a garden-path effect, since a comma is possible after the NP *the dog* but none occurs. However, no garden-path effect occurs in (332a) (or in the other sentences in (332)), and thus comma placement does somewhat depend on local attachment preferences. See also footnote 20.

that is in a $\theta$-position but currently lacks a thematic role. Thus (334b) is associated with $x_{Int}$ PLUs. Structure (334a) has no load associated with it whatsoever so the load difference between the two structures is $x_{Int} + x_{RP}$ PLUs. This difference is enough to cause a garden-path effect, as was demonstrated earlier.

### 7.3.3. Lexical Ambiguity and Recency Preference: Multiple Complement Ambiguities

Next consider the sentences in (335):

(335) a. # I convinced her children are noisy.
b. # I told the department committees will be formed next month.
c. # I showed the Canadian jewelry is expensive.
d. # I persuaded the company secretaries are valuable.

Consider (335a). This sentence is locally ambiguous over the span of the words *her children*. Two parses for the input string *I convinced her children* are given in (336):

(336) a. $[_{IP}$ $[_{NP}$ I ] $[_{VP}$ $[_{V'}$ $[_{V}$ convinced ] $[_{NP}$ her children ] $[_{CP}$ ] ]]]
b. $[_{IP}$ $[_{NP}$ I ] $[_{VP}$ $[_{V'}$ $[_{V}$ convinced ] $[_{NP}$ her ] $[_{CP}$ $[_{IP}$ $[_{NP}$ children ] ]]]]]

In structure (336a) the words *her children* correspond to the noun phrase object of *convinced*. In (336b), the word *her* is the noun phrase object of *convinced* and the word *children* is the subject of the complement clause of *convinced*. Structure (336a) is strongly preferred over (336b) and a garden-path effect results for (335).

In order to see how this garden-path effect can be predicted in this framework, consider the loads associated with each of the structures in (336). First, let us consider the Property of Recency Preference. The attachment of the NP representation of the word *children* to form (336a) requires no PRP load since the H-node that is matched is the one associated with the word *her*, the most recent word in the input string. On the other hand, the attachment of *children* as the subject of the CP complement to *convinced* involves matching an H-node associated with *convinced*, a less recent word in the input string. Thus there is a load of $x_{RP}$ PLUs associated with (336b).

Consider now the Properties of Thematic Reception and Lexical Requirement. Structure (336a) has a load of $x_{Int}$ PLUs since the verb *convinced* has a yet unassigned thematic role.[126] Structure (336b) is associated with a load of $x_{Int}$ PLUs since the noun phrase *children* requires a thematic role and does not yet receive one. Thus the total load difference between the two structures is $x_{RP}$ PLUs. Hence the garden-path effects in (335) are explained.

Now let us consider related examples in which the verb *convinced* is replaced with a

[126] This thematic role may in fact be optional. If so, then the relevant load difference increases and the garden-path effect is still predicted.

verb like *give*. Like the verb *convince*, the verb *give* subcategorizes for two complements. However, unlike the verb *convince*, neither of the complements of the verb *give* is clausal. This difference means that a structure may attach as the second complement position without being penalized for not receiving a thematic role, since it will receive a thematic role directly from the verb. Thus the load difference between attachment to the first complement and attachment as the second complement will be less than what it is for an example including a verb like *convince*. Thus a garden-path effect need not occur. In fact, a garden-path effect does not occur, as is demonstrated by the examples in (337):

(337) a. I gave her earrings on her birthday.
b. I gave her earrings to Sally.
c. John sent his company boxes for storage.
d. John sent his company boxes to the recycling plant.
e. The president brought the committee folders for their notes.
f. The president brought the committee folders to the secretary.
g. We gave the department computers on their 20th anniversary.
h. We gave the department computers to a charity.
i. I sent the American mail in order to inform him of the developments.
j. I sent the American mail to Canada.

In order to formulate an inequality representing the state of affairs in sentences like those in (337), consider the parse state after the word *earrings* is read in (337a) and (337b):

(338) a. $[_{IP}$ $[_{NP}$ I ] $[_{VP}$ $[_{V'}$ $[_{V}$ gave ] $[_{NP}$ her earrings ] $[_{PP}$ ] ]]]
b. $[_{IP}$ $[_{NP}$ I ] $[_{VP}$ $[_{V'}$ $[_{V}$ gave ] $[_{NP}$ her ] $[_{NP}$ earrings ] ]]]

The verb *give* subcategorizes for either two noun phrases or for a noun phrase and a prepositional phrase (*cf.* Kayne (1984), Larson (1988)). First consider the load associated with the structures in (338) with respect to the Property of Recency Preference. The derivation of this load is similar to that for the structures in (336). Structure (338a) is associated with no load by the PRP since the most recent H-node is matched when attaching the noun *earrings*. However, structure (338b) is associated with a load of $x_{RP}$ PLUs since a less recent H-node is matched in making this attachment. Now consider the loads associated with the structures in (338) with respect to the Properties of Thematic Reception and Lexical Requirement. Structure (338a) is associated with a load of $x_{Int}$ PLUs since a thematic role is yet unassigned by the verb *gave* in this structure. On the other hand, structure (338b) is associated with no load due either the PLR or the PTR since all thematic roles are assigned and all thematic elements receive thematic roles. Note that this is the crucial difference between (337) and (335). Thus the load difference between structures (338a) and (338b) is $x_{RP} - x_{Int}$ PLUs. Since neither (337a) nor (337b) is a garden-path sentence, I hypothesize that this load difference is not significant:

(339) $x_{RP} - x_{Int} \leq P$

## 7.4. Property Solution Space

The inequalities obtained for the loads associated with the Properties of Lexical Requirement, Thematic Reception and Recency Preference with respect to the preference constant $P$ are given in (340):

(340) a. $x_{Int} \leq P$
b. $x_{RP} > P$
c. $2x_{Int} > P$
d. $x_{RP} - x_{Int} \leq P$

This set of inequalities is consistent. It is therefore possible to solve for the possible values associated with the relevant properties in terms of the preference constant $P$. The solution space of such values is depicted in Figure 13. On this graph the load associated with the Properties of Thematic Reception and Lexical Requirement are represented as a single variable, $x_{Int}$, on the horizontal axis, while the load associated with the Property of Recency Preference, $x_{RP}$, is represented on the vertical axis. The areas that satisfy all the inequalities in (340) are then shaded on the graph, so that the intersection of all of these areas satisfies all of them. Thus every point in the shaded area of the graph in Figure 13 satisfies all the inequalities in (340). For example, the point $x_{Int} = 0.6P, x_{RP} = 1.2P$ lies inside the solution space and thus satisfies all the inequalities.

Figure 13: The Solution Space for the Inequalities in (340)

## 8. Processing Overload

In addition to providing explanations of garden-path and preferred reading effects, the Properties of Thematic Reception, Lexical Requirement and Thematic Transmission also give a plausible account of unacceptability due to processing overload.[127] Recall that it is assumed that structures associated with more than $K$ PLUs are not pursued by the human parser as a result of linguistic short term memory pressures. Thus a structure is ruled unacceptable if the load associated with that structure is greater than the maximum allowable processing load, $K$ PLUs:

(10)

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

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 that are associated with processing load,
$A_i$ is the number of times property $i$ appears in the structure.

Thus a sentence $S$ may be said to be unacceptable due to processing overload if there exists a structure $A$ in all possible parses of $S$, where $A$ represents a left substring of $S$ and the load associated with $A$ is greater that $K$ PLUs.

Although (10) dictates that structures requiring more than $K$ PLUs cannot be maintained, it does not say what happens to these overly expensive structures.[128] While there are other alternatives, the most obvious hypothesis is that these structures are simply pruned:

(341) The Node Pruning Hypothesis (NPH):
If a structure requires more processing load than the available capacity, then prune it from further consideration.

Alternatives to the Node Pruning Hypothesis consist of maintaining *part* of the overly expensive structure. See Section 8.2 for one such alternative. The choice between these hypotheses has no effect on which sentences are ruled unacceptable; the empirical differences between the hypotheses consist mainly of the location at which processing difficulty

---

[127]The Property of Recency Preference does not take part in the explanation of processing overload effects since it is necessarily a relative property: it requires the existence of an alternative structure. A crucial property of a processing overload effect is that attempts at reanalysis always fail. That is, even if the correct analysis can be explained to the subject, he/she will still have great difficulty obtaining that analysis. For example, even if a subject knows what the interpretation of a doubly center-embedded relative clause is supposed to be, he/she will still have a great deal of difficulty arriving at that interpretation. Note that garden-path sentences do not have this property: once a subject is explained the correct analysis of a garden-path sentence, the analysis can then be easily understood.

[128]Note that a structure requiring more than $K$ PLUs will be pruned independently if there exists a structure requiring $K - P$ PLUs. See Chapter 1.

is noticed. For simplicity, I will initially assume the NPH as a default hypothesis. However, none of the acceptability judgments to be discussed in this section crucially relies on the NPH; the empirical consequences of the NPH and alternative hypotheses is dealt with in Section 8.2.

This chapter is divided into four sections. The first of these sections includes syntactic data relevant to the explanation of processing overload in English. Section 8.2 discusses alternatives to the Node Pruning Hypothesis. The third section describes a small amount of processing overload data obtained from languages other than English. The final section outlines how the theory of processing overload given here might be extended into a theory of processing complexity.

## 8.1. Acceptability Effects

This section contains data relevant to the explanation of processing overload in English. The constructions that are involved in this data consist of sentential subjects, NP and sentential complements as well as a number of constructions which involve long-distance dependencies: relative clauses, *wh*-questions, clefts, pseudo-clefts, *though* preposings.[129] As in previous chapters, the judgments of processing difficulty are based on my intuitions coupled with the intuitions of many others who participated in informal acceptability judgment tasks.

### 8.1.1. Relative Clauses

Consider the unacceptable center-embedding sentences in (342) with respect to (10) and (341):

(342) 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.

The structure that results from parsing either of the sentences in (342) after the second complementizer *that* has been input is given as follows:

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

First consider (343) with respect to the Property of Thematic Reception. There are two lexical noun phrases in (343) that need thematic roles but lack them. Furthermore, there are two non-lexical NPs, operators, that are in positions that must eventually be linked to thematic roles. Thus the load associated with (343) is at least $4x_{TR}$ PLUS. Now consider (343) with respect to the Property of Lexical Requirement. Only the two complementizers have lexical requirements in (343), 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.

[129] Cowper (1976) gives a similar list of construction combinations relevant to processing overload. Her list was very useful in coming up with the set of relevant data given in Section 8.1.

Thus the total load associated with (343) is $4x_{TR} + x_{LR}$ PLUS $= 5x_{Int}$ PLUs. I hypothesize that this load is greater than the processing overload constant $K$:[130]

(345) $5x_{Int} > K$

Thus, by the Node Pruning Hypothesis, structure (343) is pruned from consideration, and since the sentences in (342) require that structure (343) be maintained in order to be successfully parsed, these sentences end up being ruled unacceptable.[131]

Note that the absence of lexical complementizers in (342b) does not change the unacceptability judgment for that construction:[132]

(347) # The man the woman the dog bit likes eats fish.

Consider one structure that results after the NP *the dog* has been input:

---

[130] While the sentences in (342) are unacceptable, sentences like hose in (344) are usually judged to be slightly better ((344a) from Kimball (1975), (344b) from Frazier & Fodor (1978)):

(344)a. ?# The guy whom the secretary we fired slept with is a real lucky dog.
b. ?# The snow that the match that the girl lit heated melted.

While these sentences are better than those in (342), they are still usually judged to be unacceptable. Following Stolz (1967) and Schlesinger (1968), I assume the improvement can be attributed to the fact that semantic/pragmatic factors are compensating for the syntactic parsing failure in (344b) and (344a). That is, it is hypothesized that although people cannot syntactically parse sentences like those in (344b) and (344a), the proper interpretations can still be calculated based on the partial parses along with consideration of the pragmatically plausible relations among the nouns and verbs. Note that no such compensation is of any help to sentences like those in (342), since all the noun phrases can plausibly fill any of the open subject or object thematic roles.

[131] See Section 8.2 for an alternative to the Node Pruning Hypothesis that predicts a different location of processing failure for sentences like those in (342).

[132] Fodor & Garrett (1967) observed that center-embedded structures that lack lexical complementizers, such as in (347), are in fact more difficult to process than the corresponding structures that contain lexical complementizers, such as in (342b). The explanation for this effect that they give appeals to the additional local ambiguity in the sentences without lexical complementizers. That is, the three initial noun phrases in (347) need not be part of a center-embedded relative clause structure; they could instead be part of a conjoined NP subject as in (346):

(346) The man the woman the dog and the cat all ate dinner together.

Fodor & Garrett hypothesize that people locally prefer the conjoined NP reading of the three initial noun phrases over the embedded relative clause reading and the additional complexity of sentences like (347) is explained.

I propose to explain the added complexity of (347) in the same way. Since (346) is grammatical and acceptable (see footnote 20 for comments regarding punctuation issues), the processing load for the initial conjunction-less NP *the man the woman the dog* must be below the maximum allowable load, by hypothesis. One way to arrive at this end would be to assume that there is a load required for the conjoined noun phrase, $x_{Int}$, and there is a load associated with maintaining a list-like phrase which requires a conjunction, but currently lacks one. Thus it will be more difficult to parse center-embedded structures with non-lexical complementizers than corresponding structures with lexical complementizers since people will be initially misled (garden-pathed) in the former case, but not in the latter case.

(348) $[_{IP}\ [_{NP}$ the man$_i$ $[_{CP}\ O_i\ [_{IP}\ [_{NP}$ the woman$_j$ $[_{CP}\ O_j\ [_{IP}\ [_{NP}$ the dog ]]]]]]]

At this point there are three lexical noun phrases, *the man*, *the woman* and *the dog* as well as two non-lexical operators, all which require thematic roles. Thus the processing load associated with this structure is $5x_{Int}$ PLUs, enough to cause processing overload.

In contrast, sentences with only one relative clause modifying a subject NP are perfectly acceptable, as is exemplified in (349):

(349) The man that Mary likes eats fish.

The processing load associated with (349) is greatest when the complementizer *that* is input. At this point in the parse, two argument NPs lacking thematic roles are present: the lexical NP *the man* and a non-lexical operator. Furthermore, the complementizer *that* requires a complement IP which is not yet present. Thus the total load associated with (349) at the point of parsing *that* is $3x_{Int}$ PLUs.[133] Since there is no difficulty in parsing (349), the inequality in (350) is obtained:

(350) $3x_{Int} \leq K$

Furthermore, when a noun phrase with two levels of center-embedded relative clauses appears post-verbally, the result is significantly more acceptable than when the NP appears pre-verbally (Eady & Fodor (1981)):

(351) 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 (351). In particular, consider the state of maximal processing load in the parse of the sentences in (351): the point of parsing the second complementizer *that*. At this point there are three NP arguments requiring thematic roles as well as a complementizer lacking its argument, so that the total load associated with this state is $4x_{Int}$ PLUs. Since each of these sentences is just acceptable, I hypothesize that this load is less than or equal to the short term memory capacity:

(352) $4x_{Int} \leq K$

Thus I assume that the maximum processing load that people can handle lies above $4x_{Int}$ PLUs but below $5x_{Int}$ PLUs.

Moreover, the ease of processing purely right or left branching structures is easily explained in the framework proposed here ((353) and (354) from Kimball (1973)):

(353) The dog saw the cat which chased the mouse into the house that Jack built.

[133] 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.

(354) My cousin's aunt's dog's tail fell off.

Consider (353). The processing load associated with the parse of this sentence starts at $x_{Int}$ PLUs when the NP *the dog* is read. When the verb *saw* is input, a thematic role is assigned to the NP *the dog*, but the total load remains at $x_{Int}$ PLUs since the verb *saw* has lexical requirements that are as yet unsatisfied. When the NP *the cat* is then input, the processing load decreases to 0 PLUs since all principles of interpretation are satisfied by the current input string. The load then goes up when the word *which* is input, marking the initiation of a relative clause. However, this load goes back down again when the next verb *chased* is read. The load decreases further when the object NP *the mouse* is read, and the cycle of minor load fluctuations repeats itself with further right-branching structure. Similarly, the processing load associated with the processing of (354) never gets to be greater than $x_{Int}$ PLUs. As a result, no difficulty is predicted with the processing of strictly left or branching structures as given in (353) and (354), as desired.

### 8.1.2. Sentential Subjects

Given the existence of the Property of Thematic Transmission in addition to the Properties of Thematic Reception and Lexical Requirement, the contrast between (355) and (356) is predicted (examples (355), (356a) and (356b) from Kimball (1973)):

(355) a. That Joe left bothered Susan.
b. That the food was good pleased me.
c. That John smokes annoys Catherine.

(356) a. # That that Joe left bothered Susan surprised Max.
b. # That for Joe to leave bothers Susan surprised Max.
c. # That for John to smoke would annoy me is obvious.

Although it is possible for a clause to be the subject of a matrix clause, as in (355), an unacceptable sentence results when the subject clause contains a further clause as its subject, as in (356). The acceptability of the sentences in (355) is explained by the Properties of Thematic Reception, Lexical Requirement and Thematic Transmission. Consider the load associated with each of the sentences in (355) after the word *that* has been processed:[134]

(357) $[_{IP}$ $[_{CP}$ that $[_{IP}$ ] ]]

The processing load associated with (357) is $2x_{Int} + x_{TT}$ PLUs since 1) the CP headed by *that* requires a thematic role, but it is currently lacking one; 2) the lexical requirements of the complementizer *that* are currently unsatisfied; and 3) this complementizer is semantically null but is in a thematic position, and its lexical requirements contain no thematic elements.

[134] I will ignore the demonstrative readings of the word *that* here. These other readings are important when considering reanalysis issues (see Section 6.9.5), but they are not relevant to the issue of absolute processability.

Since it has already been established that the load associated with the Property of Thematic Transmission $x_{TT}$ is less than the load associated with the Properties of Thematic Reception or Lexical Requirement, $x_{Int}$ (see Section 6.10), the processing load of (357) is less than $3x_{Int}$ PLUs and therefore well below the maximum allowable load. Since the load associated with the processing of the sentences in (355) is at a peak at the point of parsing the complementizer *that*, each of these sentences is acceptable.

The unacceptability of the sentences in (356) is explained by appeal to the same properties. Consider (356c). The point at which the processing load is highest occurs after the complementizer *for* has been processed:[135]

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

At this parse state there are two complementizer phrases, both which require thematic roles, but currently lack them. Furthermore, both complementizers have lexical requirements that are currently unsatisfied. That is, the complement of each complementizer neither is a confirmed node nor contains a thematic element. Finally, each of these complementizers is associated with processing load via the Property of Thematic Transmission since each is a semantically null category in a thematic position whose lexical requirements are not yet satisfied. Thus the total load associated with (358) is $4x_{Int} + 2x_{TT}$ PLUs. If we assume that this load is greater than the processing overload constant $K$, then we have an explanation of the unacceptability of the sentences in (356).

(359) $4x_{Int} + 2x_{TT} > K$

Thus the unacceptability of (356c) is explained, as desired.[136]

Furthermore, the existence of the Property of Thematic Reception predicts a subject-object asymmetry for complementizer phrases receiving thematic roles. That is, the explanation of the unacceptability of the sentences in (356) given here crucially relies on the fact that neither complementizer phrase in each of these examples receives a thematic role when it is initially input. If one were to receive a thematic role, then perhaps no processing overload effect would occur. This is, in fact, the case:

(360) a. I believe that for John to smoke would annoy me.
b. The newsman thinks that for Dan to quit politics would please the people.
c. I understand that for Fred to be at the meeting is expected.

Hence the acceptability of the sentences in (360) is explained. The maximal processing

[135]I will ignore the prepositional reading of the word *for* in these circumstances as it is not relevant to the calculation of absolute processing load.

[136]Note that if the Property of Thematic Transmission did not exist, then we would have no explanation of the processing overload effects in (356). That is, if the load associated with sentential subjects lacking thematic roles were $x_{Int}$ PLUs, then the total load associated with (358) would only be $4x_{Int}$ PLUs, not enough to cause processing overload. However, given the existence of the PTT, the processing overload effects in (356) follow immediately.

load associated with each of the sentences in (360) occurs at the point of processing the complementizer *for*. Consider the structure for (360a) at this point:

(361) $[_{IP}$ $[_{NP}$ I ] $[_{VP}$ believe $[_{CP}$ that $[_{IP}$ $[_{CP}$ for $[_{IP}$ ]] ]] ]]

The load associated with this structure $3x_{Int} + 2x_{TT}$ PLUs since 1) the CP headed by *for* is in a thematic position but currently lacks a thematic role; 2) the lexical requirements of the complementizers *that* and *for* are as yet unsatisfied; and 3) the two complementizers are both semantically null categories in thematic positions and thus are associated with load via the PTT. Crucially, the CP headed by *that* receives a thematic role in this structure and so is not associated with any load via the Property of Thematic Reception. If we assume that the load associated with this structure is less than the processing overload constant $K$, then we have an explanation for the acceptability of the sentences in (360):

(362) $3x_{Int} + 2x_{TT} \leq K$

This inequality is consistent with all of the inequalities that have been arrived at thus far.

In addition, the theory presented here provides an explanation of the fact that nominalizations of the clausal subjects in (356) are much easier to process ((363a) adapted from Kimball (1973)):

(363) a. Joe's leaving bothering Susan surprised Max.
b. John's smoking annoying me is apparent.

Since there are no complementizers in pre-verbal thematic positions in the examples in (363), the processing load associated with either never gets very high. For example, consider (363a) at the point of processing *Joe's*. No load is associated with the NP *Joe* at this point since the location of its thematic role assigner is identifiable as the head of the category which it specifies. When the gerund *leaving* is input the load increases to $x_{Int}$ PLUs since the phrase *Joe's leaving* needs a thematic role. However, no other factors contribute to load. The processing of the rest of (363a) proceeds without ever needing to sustain a large load, and thus the ease of processing these sentences is explained.

### 8.1.3. Sentential Subjects and Relative Clauses

It has been observed that two center-embedded relative clauses cause processing overload when in subject position and that a sentential subject inside a sentential subject also cause processing overload. Thus it is reasonable to test whether a relative clause inside a sentential subject or a sentential subject inside a relative clause causes processing overload. First, let us consider the case of relative clause embedded within a sentential subject ((364a) from Cowper (1976)):

(364) a. That the man who said that the book was finished didn't know his business is clear to everyone.
b. That the food that John ordered tasted good pleased him.
c. That the man that Mary loved died yesterday upset her.

Somewhat surprisingly, the sentences in (364) are acceptable.[137] In order to explain the lack of difficulty that people experience in parsing sentences like those in (364), consider the point of maximal complexity in the parse of (364b): after the second complementizer *that* has been input:

(365) $[_{IP}$ $[_{CP}$ that $[_{IP}$ $[_{NP}$ the food$_i$ $[_{CP}$ $[_{NP}$ $O_i$ ] that $[_{IP}$ ]] ]] ]]

At this point, load-bearing elements are as follows: one CP in a thematic position which currently lacks a thematic role; one lexical NP, *the food*, in a thematic position but currently lacking a thematic role; one non-lexical NP, the operator $O_i$ coindexed with *the food*; and the unsatisfied lexical requirements of the second complementizer *that*.[138] Thus the total load associated with this structure is $4x_{Int}$ PLUs. Hence the acceptability of the sentences in (364) is explained.

Now consider sentences in which sentential subjects are embedded inside relative clauses:

(366) a. # The woman that for John to smoke would annoy works in this office.
b. # The child that for the hamster to die would upset is very shy.
c. # The bureaucrats that for Iraq to invade Kuwait would affect were employed by the government.
d. # The decisions that for it to rain would influence were made by the picnic committee.

The sentences in (366) are unacceptable, providing a surprising contrast to those in (364). It turns out that the unacceptability of the sentences in (366) is explained by the theory presented here. Consider the state of parsing (366a) after the complementizer *for* has been input:

(367) $[_{IP}$ $[_{NP}$ the woman$_i$ $[_{CP}$ $O_i$ that $[_{IP}$ $[_{CP}$ for $[_{IP}$ ]]] ]]]

At this parse state 1) the NP *the soup* lacks a thematic role, 2) the operator $O_i$ lacks a thematic role, 3) the CP headed by *for* lacks a thematic role, 4) the lexical requirements of the complementizers *that* and *for* do not yet contain thematic elements and 5) the complementizer *for* is a semantically null category in a thematic position that will need to transfer a thematic role. Thus the total load associated with this structure is $5x_{Int} + x_{TT}$ PLUs,

---

[137]These facts contradict Kimball's Principle of Two Sentences (Kimball (1973)). See Section 3.6.

[138]The lexical requirements of the first occurrence of *that* are satisfied by a node containing a thematic element, the NP *the food*. Thus no load is associated with satisfying the lexical requirements of this complementizer. Furthermore, no load is associated with the Property of Thematic Transmission for either complementizer, since 1) the lexical requirements of the first complementizer have already been satisfied and 2) the second complementizer is not in a thematic position.

enough to cause processing overload.

Furthermore, we have a plausible explanation of the unacceptability of sentences like those in (368), in which the NP that is being modified by a relative clause receives its thematic role as soon as it is processed:

(368) a. # The company hired the woman that for John to smoke would annoy.
b. # Mary played with the child that for the hamster to die would upset.
c. # The government employed the bureaucrats that for Iraq to invade Kuwait would affect.
d. # The picnic committee made decisions that for it to rain would influence.

Consider (368a). The load associated with the structure that is present at the state of the parse after the complementizer *for* has been input is the same as that for structure (367) except that the NP *woman* is not associated with any load via the PTR since it already has a thematic role. Thus the total load associated with this structure is $4x_{Int} + x_{TT}$ PLUs. Since each of the sentences in (368) cause processing overload, it is reasonable to assume that this load is more than the processing overload constant $K$:

(369) $4x_{Int} + x_{TT} > K$

This inequality is consistent with all previous inequalities.

### 8.1.4. NP Complements

While multiply center-embedded relative clauses cause severe processing difficulty, embedding a relative clause inside a NP complement does not result in unacceptability ((370a) from Cowper (1976)):

(370) a. The possibility that the man who I hired is incompetent worries me.
b. The report that the armed forces that arrived first would have to stay for another year surprised me.
c. The report that aliens who knew Elvis landed in Nashville was substantiated with photographs.
d. The information that the weapons that the government built didn't work properly affected the enemy's defense strategy.
e. The assumption that all drugs that people take recreationally are harmful influences many government policy decisions.
f. The belief that all people who want to vote should be able to do so has caused many political changes over the years.

The maximal processing load associated with the processing of (370a) occurs after the word *who* is parsed. Consider the structure associated with this parse state:

(371) $[_{IP}$ $[_{NP}$ the possibility $[_{CP}$ that $[_{IP}$ $[_{NP}$ the man $[_{CP}$ $[_{NP}$ who ] $[_{C'}$ $[_{IP}$ ]]]]]]]

The load associated with this parse state is $4x_{Int}$ PLUs since: 1) the NPs *the possibility*, *the man* and *who* all require thematic roles; and 2) the lexical requirements of the non-lexical complementizer are as yet unsatisfied. Thus the acceptability of these sentences is explained.

On the other hand, when a relative clause is embedded inside an NP complement, the results are not acceptable when the head noun is in subject position ((372a) from Cowper (1976)):

(372)a. # The man who the possibility that students are dangerous frightens is nice.
b. # The woman who the report that aliens landed in Nashville pleased is a big Elvis fan.
c. # Many bureaucrats who the information that Iraq invaded Kuwait affected are employed by the government.
d. # Many decisions which the assumption that it might rain influenced were made by the picnic committee.
e. # Changes that the belief that all men are equal caused have occurred in our society.

The unacceptability of these sentences is explained by appeal to the Properties of Thematic Reception, Lexical Requirement and Thematic Transmission. Consider (372a) at the point of processing the complementizer *that*. At this point, the processing load associated with (372a) is $4x_{Int}+x_{TT}$ PLUs since: 1) the NPs *the man*, *who* and *the possibility* all require thematic roles; 2) the lexical requirements of the complementizer *that* are not yet satisfied; and 3) the complementizer *that* is a semantically null category in a thematic position that needs to transmit a thematic role. This load is enough to force processing overload, and thus the overload effects in (372) are explained as desired.

The account presented here predicts a contrast between the examples in (372) and those in (373):

(373)a. The pentagon employs many bureaucrats who(m) the information that Iraq invaded Kuwait affected.
b. The picnic committee made many decisions which the assumption that it might rain influenced.
c. There have been many changes that the belief that all men are equal has caused.

The maximal processing load reached in any of the sentences in (373) is only $3x_{Int}+x_{TT}$ PLUs, and thus these sentences are predicted to be just acceptable, as is in fact the case.[139]

[139] While the sentences in (373) are acceptable to most readers, the sentences in (374) are not as much so:

(374)a. ?# The company employs a man who the idea that war will start soon frightens.
b. ?# I saw the woman who a report that aliens landed in Nashville pleased.

The constructions in (374) are identical to those in (373) except for the fact that the verbs in the relative clauses modifying the matrix verbs in (374) are psychological verbs, while the verbs in the same positions in (373) are not. That is, extraction of the NP in object position across a subject which contains a sentential

The theory presented here correctly predicts the difficulty with the following sentences in which a sentential subject is embedded inside an NP complement:

(375) a. # Mary's belief that for John to smoke would be annoying is apparent due to her expression.

b. # The fact that for the Americans to win the war would mean many deaths is disturbing.

The maximal load associated with these sentences occurs at the point of processing the complementizer *for* in each. At this point the processing load reaches $4x_{Int} + 2x_{TT}$ PLUs, enough to cause processing overload. Furthermore, when the noun phrase in subject position in the sentences in (375) is moved to object position, acceptability greatly improves, as predicted by the theory here:

(376) a. Mary held the belief that for John to smoke would be annoying.

b. I was disturbed by the fact that for the Americans to win the war would mean many deaths.

The maximal load associated with the sentences in (376) is only $3x_{Int} + 2x_{TT}$ PLUs, not enough to cause processing overload, so the contrast between the sentences in (376) and (375) is predicted as desired.

However, the borderline acceptability/unacceptability of the sentences in (377) is unexplained under the current theory ((377d) from Cowper (1976)):

(377) a. ?# The fact that the idea that John is gay would upset his parents disturbs me.

b. ?# John's suspicion that a rumor that the election had not been run fairly was true motivated him to investigate further.

c. ?# The fact that the idea that God is everywhere is pervasive among the world's religions is not surprising.

d. ?# The possibility that the idea that students are dangerous will be abandoned is faint.

The maximum complexity associated with the sentences in (377) is $3x_{Int} + x_{TT}$ PLUs, and thus each is predicted to be acceptable by the current theory. While these sentences are acceptable to many readers, they are only just so. Many others find these sentences unacceptable. This borderline unacceptability is not predicted under the current theory. I leave it to further research to determine the cause for this unexpected result.

### 8.1.5. *Wh*-Questions

*Wh*-questions are formed in English by fronting a *wh*-constituent. While an argument *wh*-

element seems to be impossible for some independent reason. I offer no explanation of this effect here. See Belletti & Rizzi (1988) and the references cited there for more on the behavior of psych-verbs.

constituent requires a thematic role, it can be separated from its thematic role assigner by a number of intervening constituents. Consider a construction that is acceptable under the theory of processing overload presented here, but only just so. For example, consider a construction whose maximal load is $4x_{Int}$ PLUs. Preposing a *wh*-constituent before this construction creates a combination of constructions that the theory predicts to be unacceptable. Whether or not such predictions are borne out is a strong test of the given theory.

For example, consider relative clause formation. A single relative clause modifying a subject NP is acceptable, as is a *wh*-question whose subject contains a single relative clause modifying a subject NP:[140]

(379) a. The man that Mary likes eats fish.
b. What did the man that Mary likes eat?

The acceptability of (379b) is as predicted by the theory given here. The maximal processing load associated with (379a) is $3x_{Int}$ PLUs. Thus the maximal load associated with (379b) is $4x_{Int}$ PLUs since an additional NP requiring a thematic role is present at the point of maximal load in this sentence. Since $4x_{Int}$ PLUs is not enough to cause processing overload, (379b) is correctly predicted to be acceptable.

Consider now the sentences in (380) whose maximal processing load is $4x_{Int}$ PLUs:[141]

(380) a. John donated the furniture that the repairman that the dog bit found in the basement to charity.
b. The salesman distributed the leaflets that the secretary that was on her lunch break printed to many companies.
c. The wealthy businessman contributed the profit that the company that he owned made in the last quarter to the United Way.

The maximum load associated with these sentences is $4x_{Int}$ PLUs at the point of processing the second instance of the complementizer *that* in each (see the sentences in (351) for similar examples). The unacceptability of the following sentences therefore comes as no surprise:

[140]While embedding relative clauses inside *wh*-questions gives a good test of the theory of processing overload, embedding sentential subjects inside *wh*-questions does not give as good a test, since it turns out that simple interrogatives which include sentential subjects are bad for some reason (Koster (1978)):

(378) * Did that John smokes annoy Mary ?

I assume that the unacceptability of (378) is due to the grammar rather than to processing overload.

[141]While (379b) is perfectly acceptable, the sentences in (380) (along with those in (351)) are only just acceptable, clearly more difficult than (379b). However, the theory presented here predicts that these sentences should all be equally easy (hard) to process. This processing complexity difference is unexplained by the properties and loads presented here. I leave the explanation of this difference to future work.

(381) a. # Who did John donate the furniture that the repairman that the dog bit found (in the basement) to?

b. # Who did the salesman distribute the leaflets that the secretary that works efficiently printed (on the office laser printer) to?

c. # Who did the wealthy businessman contribute the profit that the company that he owned made (in the last quarter) to?

The maximal load associated with each of these sentences is $5x_{Int}$ PLUs: $4x_{Int}$ PLUs for the doubly center-embedded relative clause in object position plus $x_{Int}$ PLUs for the fronted *wh*-constituent. Hence the unacceptability of these sentences is explained.

Additional support of the processing overload theory given here is given by the acceptability of the sentences in (382):

(382) a. Who did the information that Iraq invaded Kuwait affect most?

b. What decisions did the assumption that it would rain tomorrow influence?

c. What changes did the belief that all men are equal cause?

The maximal processing load associated with each of these sentences is $3x_{Int}+x_{TT}$ PLUs, and thus each is correctly predicted to be acceptable.

Furthermore, recall that the maximal processing load associated with each of the following acceptable sentences is $4x_{Int}$ PLUs:

(370d) The information that the weapons that the government built didn't work properly affected the enemy's defense strategy.
(370e) The assumption that all drugs that people take recreationally are harmful influences many government policy decisions.
(370f) The belief that all people who want to vote should be able to do so has caused many political changes over the years.

Now consider sentences formed by questioning the matrix objects in the sentences in (370):

(383) a. # Who does the information that the weapons that the government built don't work properly affect most?

b. # What decisions does the assumption that all drugs that people take recreationally are harmful influence?

c. # What changes has the belief that all people who want to vote should be able to do so caused?

The unacceptability of these sentences follows from the fact that the maximal processing load in each is $5x_{Int}$ PLUs, a load which has been shown to be greater than the processing overload constant $K$ PLUs.[142]

[142] As noted in footnote 139, extraction of the object of a psych-verb across a subject containing a sentential element seems to be unacceptable independently. Thus the following sentences are also unacceptable:

Finally, the extreme unacceptability of (385b) relative to (385a) is also predicted by the current theory:

(385) a. # The woman who the man who hired me married dislikes Fred.
b. # Who does the woman who the man who hired me married dislike ?

The maximal load associated with (385a) is $5x_{Int}$ PLUs, while the maximal load associated with (385b) is $6x_{Int}$ PLUs and thus the distinction between the two sentences is predicted.

### 8.1.6. Clefts

The cleft construction is another construction that allows a single constituent lacking a thematic role to be placed before structures which require large processing loads, thus further testing the theory of processing overload presented here. First consider the sentences in (386):

(386) a. The man that Ellen married saw a fish on the highway.
b. It was a fish that the man that Ellen married saw on the highway.

The maximal processing load associated with (386a) is $3x_{Int}$ PLUs. Thus the maximal processing load associated with (386b) is $4x_{Int}$ PLUs since this sentence includes an additional NP locally lacking a thematic role, the NP *a fish*. Sentence (386b) is acceptable, as predicted.

Furthermore, the sentences in (387) are also correctly predicted to be acceptable:

(387) a. It was the Americans that the information that Iraq invaded Kuwait affected most.
b. It was the jury that that the rumor that the accused man was guilty influenced.
c. It was political change that the belief that all people should be able to vote caused.

The maximal load in each of the above sentences is $3x_{Int} + x_{TT}$ PLUs, and hence the acceptability of these sentences sentences is explained.

Consider now the contrast between sentences (370a) and (370b) on the one hand and the sentences in (388) on the other:

(370d) The information that the weapons that the government built didn't work properly affected the enemy's defense strategy.
(370e) The assumption that all drugs that people take recreationally are harmful influences many government policy decisions.
(370f) The belief that all people who want to vote should be able to do so has caused many political changes over the years.

---

(384)a. # Who does the possibility that the man who I hired is incompetent worry?
b. # Who did the report that the armed forces that arrived first would have to stay for another year surprise?

(388) a. # It was the enemy's defense strategy that the information that the weapons that the government built didn't work properly affected.

b. # It is government policy decisions that the assumption that all drugs that people take recreationally are harmful influences.

c. # It is political change that the belief that all people who want to vote should be able to do so has caused.

The point of maximal complexity in sentences (370a) and (370b) is $4x_{Int}$ PLUs. Thus the maximal load associated with the sentences in (388) is $5x_{Int}$ PLUs, and thus their unacceptability is explained.

Finally consider the contrast between the sentences in (389) and (390):

(389) a. For the weapons to work properly would affect the enemy's defense strategy.

b. For all recreational drugs to be proven harmful would influence government policy decisions.

(390) a. # It is the enemy's defense strategy that for the weapons to work properly would affect.

b. # It is government policy decisions that for all recreational drugs to be proven harmful would influence.

The sentences in (389) contain sentential subjects. The sentences in (390) are clefted versions of those in (389). While the sentences in (389) are acceptable, those in (390) are not. This unacceptability is explained by the fact that the maximal load in the sentences in (390) is $4x_{Int} + x_{TT}$ PLUs, enough to cause processing overload.

### 8.1.7. Pseudo-Clefts

Pseudo-clefts provide another test for the theory presented here. Consider the sentences in (391):

(391) a. The woman that John married likes smoked salmon.

b. What the woman that John married likes is smoked salmon.

I assume a free relative analysis of pseudo-clefts so that the structure for (391b) is as follows:

(392) $[_{IP}$ $[_{CP}$ $[_{NP}$ what$_i$ ] $[_{IP}$ $[_{NP}$ the woman that John married ] $[_{VP}$ likes $t_i$ ]]] $[_{VP}$ is fish ]]

The maximal processing load associated with (391a) is $3x_{Int}$ PLUs. Thus the maximal processing load associated with (391b) is $4x_{Int}$ PLUs, since in this sentence there is one additional argument requiring a thematic role, the *wh*-word *what*. Thus the acceptability of (391b) is predicted, as desired.

The acceptability judgments of the following sets of sentences are also explained under the theory developed here in a manner similar to that for previous sections.

(393) a. What the rumor that the accused man had robbed a bank influenced was the judge's decision.

b. Who(m) the information that Iraq invaded Kuwait affected most were the Americans.

(394) a. # What the information that the weapons that the government built didn't work properly affected was the enemy's defense strategy.

b. # What the assumption that all drugs that people take recreationally are harmful influences is government policy decisions.

(395) a. # What for the weapons to work properly would affect is the enemy's defense strategy.

b. # What for all recreational drugs to be proven harmful would influence is government policy decisions.

### 8.1.8. *Though*-Preposing

*Though*-preposing provides a test for the theory much like that provided by *wh*-questions, clefts and pseudo-clefts. Consider the sentences in (396):

(396) a. The man that Ellen married is intelligent.

b. Intelligent though the man that Ellen married is, he has no sense of humor.

The maximal processing load associated with (396a) is $3x_{Int}$ PLUs and thus the maximal load associated with (396b) is $4x_{Int}$ PLUs since it contains an additional displaced argument, the adjective phrase *intelligent*. Hence (396b) is correctly predicted to be acceptable.

The appropriate contrasts in the following sets of sentences are also predicted by the theory given here in much the same way that the theory predicts the acceptability effects in the two previous sections:[143]

---

[143] *Tough*-movement might seem to provide another test of the theory offered here, in much the same way that *wh*-questions, clefts, pseudo-clefts and *though*-preposing examples do. The sentences in (398) are sentences in which the matrix object in each of the sentences in (397) is displaced via *tough*-movement to the matrix subject position:

(397) a. It is easy for Mary to please John.

b. The man that Ellen married is intelligent.

(398) a. John is easy for Mary to please.

b. The boss is tough for the man that Ellen married to please.

However, many English speakers find the sentences in (398) to be only marginally grammatical. As a result, acceptability judgments of more complex *tough*-movement sentences can be more difficult. Thus I leave sentences involving *tough*-movement to future research.

(399) a. Shocking though the news that Iraq invaded Kuwait was, even worse news was yet to come.

b. True though the rumor that the accused man had once robbed a bank was, this information was irrelevant to the present case.

c. Depressing though the belief that one race is superior to others is, this belief still persists in many parts of the world.

(400) a. The information that the weapons that the government built didn't work properly was surprising.

b. The assumption that all drugs that people take recreationally are harmful is influential in the government's decisions.

(401) a. # Surprising though the information that the weapons that the government built didn't work properly was, no one took advantage of the mistakes.

b. # Influential though the assumption that all drugs that people take recreationally are harmful is in the government's decisions, this assumption may change with time.

(402) a. For the weapons to work properly would be surprising for the general populace.

b. For all recreational drugs to be proven harmful would be influential in government policy decisions.

(403) a. # Surprising though for the weapons to work properly would be for the general populace, it would not surprise some military officials.

b. # Influential though for all recreational drugs to be proven harmful would be in government policy decisions, such an assertion has not yet been demonstrated.

### 8.1.9. Property Solution Space

The inequalities obtained for the loads associated with the Properties of Lexical Requirement, Thematic Reception and Thematic Transmission with respect to the processing overload constant $K$ are given in (404), where the loads associated with the Properties of Thematic Reception and Lexical Requirement are represented as a single variable, $x_{Int}$.

(404) a. $5x_{Int} > K$

b. $4x_{Int} \leq K$

c. $4x_{Int} + x_{TT} > K$

d. $3x_{Int} + 2x_{TT} \leq K$

This set of inequalities is consistent. The solution space is depicted in Figure 14. On this graph the load associated with the Properties of Thematic Reception and Lexical Requirement, $x_{Int}$, is depicted on the horizontal axis, while the load associated with the Property of Thematic Transmission, $x_{TT}$, is represented on the vertical axis. The areas that satisfy all the inequalities in (404) are shaded on the graph, so that the intersection of all of these areas satisfies all of them. Note that if $x_{TT} = x_{Int}/2$ as suggested in Section 6.10 then all of the overload inequalities are satisfied. In particular, inequality (404c) which states that $4x_{Int} + 2x_{TT} > K$ then reduces to inequality (404a) which states that $5x_{Int} > K$. Furthermore, inequality (404d) reduces to inequality (404b) which states that $4x_{Int} \leq K$.

Figure 14: The Solution Space for the Inequalities in (404)

See Chapter 10 for demonstrations that all the inequalities that are obtained in this thesis are consistent.

## 8.2. The Node Pruning Hypothesis and the Least Recent Nodes Hypothesis

The theory presented here makes a reasonable division between acceptable and unacceptable sentences by the use of syntactic (thematic) properties. In the prediction of unacceptability, it has been assumed that structures which require more than the maximum available memory capacity, $K$ PLUs, are simply pruned from consideration by the parser. This assumption is given in (405):

(405) The Node Pruning Hypothesis (NPH):
If a structure requires more processing load than the available capacity, then prune it from further consideration.

For example, in the processing of a doubly center-embedded sentence like (406), the structure in (407) is required for a successful parse, but this structure is deleted due to the excessive processing load that it requires.

(406) # The man the woman the dog bit likes eats fish.

(407) $[_{IP}$ $[_{NP}$ the man$_i$ $[_{CP}$ $O_i$ $[_{IP}$ $[_{NP}$ the woman$_j$ $[_{CP}$ $O_j$ $[_{IP}$ $[_{NP}$ the dog ]]]]]]]]

As a result, the parse eventually fails. However, it need not be the case that exceedingly complex structures are simply pruned from consideration; instead of removing the entire structure, it is possible that only a portion of that structure is removed. Frazier (1985) suggests such a heuristic of partial structure removal, the Disappearing Syntactic Nodes Hypothesis:

(408) The Disappearing Syntactic Nodes Hypothesis:
At points of high complexity, syntactic nodes will tend to be forgotten if they dominate no lexical material or only dominate material that has already been semantically combined.

Frazier's hypothesis relies heavily upon her own definition of complexity as well as her own notion of syntactic/semantic interaction, which are quite different from those assumed here. As a result, (408) cannot easily be reformulated in this framework. However, other heuristics of structure pruning can be formulated. One obvious candidate for such a heuristic is given in (409), which is clearly related to the Property of Recency Preference (see Chapter 7):

(409) The Least Recent Nodes Hypothesis (LRNH):
If a structure requires more processing load than the available capacity, then selectively remove (forget) nodes directly dependent on the least recent words in the input string until the load associated with the structure is lowered below the desired threshold.

Note that both (405) and (409) make the same predictions with respect to the unacceptability judgments of all the sentences given so far in this section. However, the two hypotheses make different predictions with respect to the point at which processing difficulty will be noticed. For example, consider (410) with respect to both the Node Pruning Hypothesis and the Least Recent Nodes Hypothesis:

(410) # The man that the woman that the dog bit likes eats fish.

When the second instance of the word *that* is input, it is necessary to build a structure that requires more memory than the limited capacity of short term memory allows, structure (411):

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

The Node Pruning Hypothesis predicts that people will now prune this structure from consideration. Since there are no other structures for the input, parsing should stop at this point. Thus people should notice the breakdown in the processing of sentence (410) after reading the second instance of the word *that*.

On the other hand, the Least Recent Nodes Hypothesis predicts that people will alter structure (411) by discarding the partial structures directly dependent on the least recent words in the input until the load associated with the surviving structure is less than the maximum available. Thus the NP structure headed by *the man* is discarded. Furthermore, the CP modifying this NP is also discarded, since its existence in this structure is dependent on the existence of the head NP.[144] Hence structure (412) remains:

(412) $[_{IP}\ [_{NP}$ the woman$_j$ $[_{CP}\ O_j$ that $[_{IP}$ ]]]]

The load associated with this structure is below the maximum allowable load, so that processing on (410) now continues. All proceeds without difficulty until the verb *eats* is input. At this point, the structure for the input will be as in (413):

(413) $[_{IP}\ [_{NP}$ the woman$_j$ $[_{CP}\ O_j$ that $[_{IP}\ [_{NP}$ the dog ] bit ]]] saw ]

The verb *eats* cannot attach anywhere in this structure, and parsing fails. Thus although both the Node Pruning Hypothesis and the Least Recent Nodes Hypothesis predict that a sentence like (410) is unacceptable, the two hypotheses make different predictions as to where people will first perceive difficulty.

The intuition shared by most people is that the point of breakdown in doubly center-embedded structures occurs somewhere after the first verb has been input, either in the second or third verb of that structure. In a pilot experiment performed by Howard Kurtzman (personal communication), subjects were to press a button to bring on each new word in a multiply center-embedded relative clause sentence and stop when they felt that they had stopped understanding the sentence.[145] The results of this experiment seemed to indicate that people broke down on doubly center-embedded structures like (410) when reading the second verb. In addition, most of Kurtzman's subjects reported informal judgments that failure occurred on the second verb.

In contrast, Frazier (1985) suggests that sentences like (414a) are often accepted as grammatical, while sentences like (414b) are rejected ((414a) from Frazier (1985); (414b) adapted from Frazier (1985)):

(414) a. * The patient the nurse the clinic had hired met Jack.
b. # The patient the nurse the clinic had hired admitted met Jack.

The implication of these judgments is that, under Frazier's view, people break down when reading the third verb in a doubly center-embedded sentence like (414b), not the second as is suggested by Kurtzman's preliminary study.

Although the data to be accounted for is not consistent, there is agreement among the

[144] That is, the operator in the specifier position of this CP has been coindexed with the NP *the man*. Without the head NP, this structure can no longer exist.

[145] The sentences in Kurtzman's study all contained the overt complementizer *that* in all complementizer positions.

preliminary study and intuitions that the point of processing breakdown in a doubly center-embedded sentence occurs somewhere after the first verb is input. As noted earlier, the LRNH predicts that people will break down when reading the third verb in doubly center-embedded sentences like (410) and (414b). The LRNH also predicts that people will not notice the ungrammaticality of a sentence like (414a). Thus the LRNH makes predictions that agree with Frazier's intuitions, but disagree with Kurtzman's preliminary study.

The Node Pruning Hypothesis, on the other hand, predicts that people will experience processing difficulty before reading the first verb in a doubly center-embedded sentence. Thus, by itself, the NPH seems to make an undesirable prediction with respect to the processing of doubly center-embedded structures. However, a number of complicating factors must be considered before ruling out the NPH. Consider once again (406) and the associated center-embedded structure associated with it at the point of processing the noun *dog*:

(406) # The man the woman the dog bit likes eats fish.

(407) $[_{IP}$ $[_{NP}$ the man$_i$ $[_{CP}$ $O_i$ $[_{IP}$ $[_{NP}$ the woman$_j$ $[_{CP}$ $O_j$ $[_{IP}$ $[_{NP}$ the dog ]]]]]]]]

Since the load associated with (407) is greater than that which is allowable in memory at one time, it might seem that the Node Pruning Hypothesis would predict that processing should stop at this point in the parse, thus contradicting the intuition that people don't experience processing failure until the second or third verb. However, the fact that structure (407) is pruned does not mean processing stops here (Fodor & Garrett (1967), *cf.* footnote 132). That is, processing may continue as long as there are other acceptable structures for the input string, ones that may not in fact be useful for the parse of the entire input string. In particular, the input *the man the woman the dog* can be continued as in (415):

(415) The man the woman the dog and the cat all ate dinner together.

Since a structure for the conjoined NP reading of the initial noun phrases can be maintained, people do not notice processing breakdown until some point after the processing of the word *dog* in (406). Furthermore, the input string *the man the woman the dog* can be continued as in (418), to give an ungrammatical, but still comprehensible sentence:[146]

[146] The fact that a sentence is ungrammatical and has no possible grammatical continuations does not mean that the sentence is necessarily difficult to interpret (Katz (1964)). For example, the sentences in (416) are ungrammatical but not difficult to understand ((416a) from Davidson (1967); (416b) from Freedman & Forster (1985)):

(416)a. * The child seems sleeping.
b. * The boys was eating their dinner.

While people recognize that the sentences in (416) are ungrammatical, there is no difficulty understanding them, since each sentence differs only slightly from a grammatical one. That is, a sentence like (416a) is easily corrected to *the child seems to be sleeping* or *the child seems to sleep* and (416b) is easily recognized as an ungrammatical variant of *The boys were eating their dinner* (Crain & Fodor (1985), *cf.* Freedman & Forster (1985)). However, it is currently unclear just what differentiates correctable ungrammaticalities, as in (416), from uncorrectable ungrammaticalities, such as in (417) (from Chomsky (1957)):

(417) * Furiously sleep ideas green colorless.

(418) The man the woman the dog ate dinner together.

Since there seems to be little difficulty with (418), which contains three NPs followed by a verb phrase, it is possible that people parse doubly center-embedded sentences like (406) without noticeable difficulty until the second verb appears, *e.g.*, *likes* in (406). Thus a possible explanation for processing breakdown at the second verb or beyond in doubly center-embedded sentences with non-lexical complementizers is given by the ability of the human processor to maintain other readings at the point of highest complexity. While the processing load associated with a conjunctive NP missing a conjunction will be high owing to the fact that it represents an ungrammatical structure, it can still be processed.

Furthermore, it is possible that the human processor is maintaining other inexpensive structures which represent ungrammatical but acceptable sentences. If this is so, then the Node Pruning Hypothesis might still be correct. For example, a possible alternative structure for the input *the man that the woman that the dog* consists of the matrix subject *the man* modified by a single relative clause, whose subject is a conjunction of the NPs *the woman* and *the dog*:

(419) $[_{IP}$ $[_{NP}$ the man $[_{CP}$ that $O_i$ $[_{IP}$ $[_{NP}$ $[_{NP}$ the woman ] $[_{NP}$ the dog ] ]]]]

This structure is not in fact a valid representation of the input, since the second instance of the complementizer *that* has been ignored. However, it seems that people may still build something like (419), in spite of the fact that it does not accurately represent the input string.[147] Evidence for this hypothesis comes from a study by Blumenthal (1966). In this study subjects were asked to paraphrase a number of center-embedded sentences in a way such that the sentences would be easier to understand. Most of the subjects then paraphrased the sentences as if they were singly center-embedded structures with either a conjunction of relative clauses modifying the matrix subject or a single relative clause with a conjoined subject and conjoined verb. If a structure like (419) can be built for the initial input of (410), then recognition of processing breakdown will not occur until the second or third verb is encountered. Thus if the human parser builds structures like that in (419), the Node Pruning Hypothesis can be maintained. As a result the choice between the NPH and the LRNH depends strongly on what structures people build for ungrammatical input. Since not much is known in this area, I leave this issue until more conclusive empirical evidence

---

One obvious difference between the sentences in (416) and the string of words in (417) is that the sentences in (416) each contain only a single violation, plausibly an agreement violation in both cases. In order to account for the acceptability of the sentences in (416), it is hypothesized that when no attachments are possible in the parse of an input string, agreement processes are relaxed (*cf.* Crain & Fodor (1985)). Hence the NP *the boys* will be permitted to attach as subject of the auxiliary verb *was* in (416b) (possibly with correction to the verb). Similarly the present participle verb *sleeping* will be permitted to attach as the complement of the verb *seems*, which normally takes an infinitival complement. Thus an interpretation for the sentences in (416) will be possible.

While the intuition that sentences like those in (416) are not difficult to process is plausible, I know of no on-line psycholinguistic studies confirming this intuition; it is therefore left to future research to resolve this issue. See, however, Mistler-Lachman (1975), Moore & Biederman (1979) and the references cited in each for evidence regarding the processing of certain kinds of ungrammatical input.

[147] See Bever (1970) for a similar argument.

can be obtained.

### 8.3. Cross-Linguistic Predictions

The examples discussed thus far have all been English ones. A strong test of the theory presented here is given by data from other languages. This section contains a small sample of cross-linguistic data that seems to corroborate the theory given here.[148] However, the data set which is considered is very small, so that the results can only be seen as suggestive. Clearly much more cross-linguistic data must be considered before any strong conclusions can be made.[149]

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 (420) is unacceptable, as expected:

(420) # 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 $\theta$-role-requiring constituents should be acceptable. This is in fact the case, as the German example (421) illustrates:

(421)
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 (421) is predicted to acceptable, as desired.

Another good test for the theory presented here comes from cross-serial dependencies in Dutch.[150] In examples of cross-serial dependency, noun phrase arguments appear at the beginning of a subordinate clause, followed by their thematic role assigning verbs. As predicted,

[148] I would like to thank my informants: Alex Franz, Hiroaki Kitano, Ingrid Meyer, Teruko Mitamura and Michel Nederlof.

[149] In fact, little cross-linguistic data has been examined as yet with respect to local preference effects such as garden-path effects. Clearly further work in this area must also be undertaken.

[150] 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 (422):

(422)
... dat Jan Piet de kinderen zag helpen zwemmen.
... that Jan Piet the children saw help swim
"... that Jan saw Piet help the children swim."

The addition of a further NP argument gives a construction which is on the border of being acceptable:

(423)
?#... 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."

A fifth NP argument causes complete unacceptability:

(424)
# ... 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 pattern of results is as predicted by the framework presented here. Five NP arguments locally lacking thematic roles force a load of $5x_{Int}$, too much for human short term memory capacity.[151]

Evidence from the processing of Japanese also seems to support 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 (425):

(425)
Jon wa Fred ga Biru o sukida to omotteiru.
John TOPIC Fred NOM Bill ACC likes COMP thinks
"John thinks that Fred likes Bill"

Sentence (425) is perfectly acceptable, as is predicted by the theory presented here. The

[151]While these results are suggestive, it turns out that the theory of processing difficulty does not account for the results of work done by Bach, Brown & Marslen-Wilson's (1986). In their experiments Bach *et al* found that cross-serial structures in Dutch were easier to process than the corresponding center-embedded structures in German. In particular, given the same number of arguments missing thematic roles, German processing is more difficult than Dutch processing when comparing cross-serial and center-embedded structures in the respective languages. Hence the properties presented here cannot account for the difference between the languages. I leave it to future work to determine what properties or parser operations could explain this difference inside the current framework. See Joshi (1990) for an account of this difference in terms of embedded pushdown automata.

processing load associated with (425) 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.

Furthermore, the presence of four $\theta$-role-requiring arguments causes processing overload:

(426)
# Jon wa Mary ga Fred ga Biru o sukida to sinziteiru to omotteiru.
John TOPIC Mary NOM Fred NOM Bill ACC likes COMP believes COMP thinks
"John thinks that Mary believes that Fred likes Bill."

The maximal processing load associated with the processing of (426) under the assumptions given here occurs at the fourth NP *Biru*. At this point the load is at least $4x_{Int}$ PLUs since there are four NP arguments all requiring thematic roles. This sentence causes a processing overload effect, which is slightly surprising given that more than $4x_{Int}$ PLUs are required to cause processing overload in English. However, other properties of syntactic structure, such as the Property of Thematic Transmission, may also be in effect in examples like (426). I will leave this issue until it is known more precisely what structures the human linguistic processor builds in the parse of SOV structures like (426).[152]

An additional embedding causes even more striking unacceptability:

(428)
# Jon wa Mary ga Fred ga Sam ga Biru o sukida to omotteiru to sinziteiru 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."

Thus the theory of processing overload given here seems to make reasonable initial predictions regarding the processing of Japanese. Much future work is still needed, however.

[152]While (426) is unacceptable, (427), which is also initiated by four arguments, is acceptable (Brad Pritchett, personal communication):

(427)
Jon wa Fred ga biiru o Dave ni ageta koto o kiita.
John TOPIC Fred NOM beer ACC Dave DAT gave COMP ACC heard
"John heard that Fred gave beer to Dave"

The load associated with (427) at the point of processing the NP *Dave* is also at least $4x_{Int}$ since four arguments require thematic roles at this point. Thus the Property of Thematic Reception does not distinguish the acceptability difference between (427) and (426). I hypothesize that other structural properties will make this differentiation possible (*e.g.* the Property of Thematic Transmission). I leave this issue for future work.

## 9. Non-syntactic Effects

While syntactic effects are significant in the processing of natural language, numerous other factors are at least as important. It is hypothesized that there exist lexical, semantic, pragmatic and discourse-level properties which also require significant processing loads. Exactly how these non-syntactic factors interact with the syntactic ones has been a topic of major discussion for many years. This debate is surveyed in Section 9.1. A number of non-syntactic factors are surveyed in the remainder of this chapter.

### 9.1. Modularity

There exists a longstanding debate in the psycholinguistic literature as to whether or not there is interaction among lexical, syntactic, semantic, pragmatic and contextual processes. Forster (1979) hypothesized that lexical, syntactic and message-level (semantic, pragmatic and contextual) processes are serially-linked and autonomous. That is, when words are input to the linguistic processor, they are first fed to the lexical processor, whose output is fed to the syntactic processor, whose output is received by the message-level processor, which produces the final interpretation of the input sentence. Crucially, there is no feedback from the syntactic processor to the lexical processor, nor is there any feedback from the message-level processor to either the lexical or syntactic processors. A second crucial assumption of some autonomous models is the serial hypothesis: that each processor can work on at most one representation at a time (Frazier & Rayner (1982), Rayner, Carlson & Frazier (1983), Ferreira & Clifton (1986)). While this model is desirable on theoretical grounds since it makes strong, falsifiable predictions, much empirical data seems to contradict it (see, for example, Tyler & Marslen-Wilson (1977), Marslen-Wilson & Tyler (1980, 1987), Kurtzman (1985), Crain & Steedman (1985), Taraban & McClelland (1988, 1990), Altmann (1988), Altmann & Steedman (1988), Ni & Crain (1989), Stowe (1989) and Trueswell, Tanenhaus & Garnsey (submitted) among others). While the serial autonomy hypothesis predicts backtracking whenever the lexical or syntactic preferences are over-ridden by semantics, pragmatics or context, these studies demonstrate there are circumstances in which no such backtracking takes place.

It turns out that the empirical difficulties with the autonomy model can be partially alleviated by dropping the serial hypothesis (Forster (1979), Crain & Steedman (1985)). That is, if more than one structure can be evaluated at a time by each of the subprocessors, then backtracking need not occur in order to get a preference for semantically, pragmatically or contextually more appropriate readings. Of course, while parallelism allows some structures to be built earlier than would be possible in a serial model, unconstrained parallelism is not psychologically plausible because of the existence of garden-path and other related effects. Thus the parallelism must be constrained in some way. One possibility that maintains Forster's autonomy hypothesis is to apply constraints inside each module. This possibility seems to be empirically disconfirmed, however, by the existence of evidence showing that message-level effects can over-ride syntactic preferences on-line (see, for example, Kurtzman (1985), Taraban & McClelland (1988, 1990), Stowe (1989) and Trueswell *et al* (submitted)).

As a result, some degree of interaction is necessary.

Crain & Steedman (1985) distinguish two types of interaction: *weak* and *strong* interaction. A strongly interactive model allows results of any subprocessor to dictate what structures are to be built in another subprocessor. Thus in a strongly interactive model the message-level processor can dictate what syntactic structures to build. On the other hand, a weakly interactive model assumes the same serial dependence of lexical, syntactic and message-level processors that is assumed in Forster's autonomous model, with one difference: the results of a later subprocessor may be used to locally eliminate structures from consideration by earlier subprocessors. Thus the lexical and syntactic processors may propose structures that the message-level processor rules out on-line, so that no further syntactic analysis is performed on these structures.

While much evidence seems to contradict non-interactive models, currently I know of no empirical evidence that rules out either the strongly or weakly interactive hypothesis. Thus I make no assumption regarding this issue here.[153]

## 9.2. Lexical Preferences

Consider the sentences in (429) (from Ford, Bresnan & Kaplan (1982)):

(429) a. The woman wanted the dress on that rack.
b. The woman positioned the dress on that rack.

There are two possible attachment sites for the prepositional phrase *on that rack* in each of the sentences in (429): the NP *the dress* and the matrix verb (*wanted* or *positioned*). The preferred interpretation of (429a) links the PP to the NP *the dress*, while the preferred interpretation of (429b) links the PP to the matrix verb *positioned*. The prospective matrix verb attachments are given in (430):

(430) a. $[_{IP}$ $[_{NP}$ the woman $[_{VP}$ wanted $[_{PP}$ $[_{NP}$ the dress ] $[_{PP}$ on that rack ]]]]
b. $[_{IP}$ $[_{NP}$ the woman $[_{VP}$ positioned $[_{NP}$ the dress ] $[_{PP}$ on that rack ]]]

In (430a) the verb *wanted* has only one complement: a small clause prepositional phrase (Stowell (1981)). In (430b) the verb *positioned* has two complements: a THEME NP *the dress* and a GOAL PP *on that rack*. While these structures are different, the PP *on that rack* attaches into thematic positions in both representations. Since (430b) is the preferred structure for (429b) and (430a) is not the preferred structure for (429a), there can be no explanation of the preferred interpretations for the sentences in (429) based only on complement/non-complement syntactic preferences. As a result of similar arguments, Ford, Bresnan & Kaplan conclude that an effect of *lexical preference* is taking place here. That is, there are at least two lexical entries for the word *want*, one which subcategorizes for an NP and one which

[153]See Fodor (1983) for philosophical arguments in favor of modularity in language processing.

subcategorizes for a PP small clause.[154] It is hypothesized that the NP subcategorization is ranked higher so that the NP modifier interpretation of the PP *on that rack* is the preferred structure for (429a).

I assume a similar analysis inside the framework presented here. In particular, I assume that a word has higher processing loads associated with its less preferred readings relative to the preferred readings. Thus the most highly preferred reading of a word will be associated with the lowest processing load for that word. Hence there will be a very low processing load associated with the reading of *want* that subcategorizes for an NP, but a relatively high processing load associated with the reading of *want* that subcategorizes for a prepositional phrase small clause. As a result of these processing load rankings, the NP reading of *the dress on that rack* is preferred over the PP small clause reading of the same input string in (429a). Only one lexical entry for the verb *position* is necessary to parse the two possible structures for (429b): one that takes an optional GOAL thematic role. The preference in (429b) for the matrix verb interpretation is explained by the difference in processing load between the two possible syntactic structures. See Section 6.3 for further details on this analysis.

The garden-path effect in (431) (from Ford, Bresnan & Kaplan (1982)) is also explained by appeal to lexical preferences:

(431) # The boy got fat melted.

This sentence is locally ambiguous at the point of parsing the word *fat*. There are two structures at this parse state: one in which an adjectival reading of the word *fat* is attached as complement of the verb *got*, and another in which a nominal reading of *fat* is attached as subject of a small clause verb phrase complement of *got*:

(432) a. $[_{IP}$ $[_{NP}$ the boy ] $[_{VP}$ got $[_{AP}$ fat ]]]
b. $[_{IP}$ $[_{NP}$ the boy ] $[_{VP}$ got $[_{VP}$ $[_{NP}$ fat ]]]]

Structure (432a) is strongly preferred at this point in the parse, and a garden-path effect results since it is structure (432b) that is necessary for a grammatical parse of (431). The explanation of the strong local preference in the parse of (431) is given by lexical preference: the adjectival complement reading of the verb *got* is preferred to the verbal small clause complement reading and the adjectival reading of *fat* is preferred to the nominal reading. Structure (432b) contains both of the more expensive readings of the words *got* and *fat* so that there is a high processing load associated with this structure. The processing load associated with (432a) is much lower and is therefore preferred.

The garden-path effects in (433) can be explained similarly:

(433) # The old man the boats.

[154]Note that the lexical entries differ primarily on subcategorization information: they have the same category and similar semantic structure. This similarity, of course, need not exist. The two entries could be entirely independent.

There are two possible structures for the input string *the old man*:

(434) a. $[_{NP}$ $[_{DP}$ the ] $[_{AP}$ old ] man ]
b. $[_{IP}$ $[_{NP}$ the old ] $[_{VP}$ man ]]

In structure (434a) the adjectival reading of *old* is attached to the nominal reading of *man*. Structure (434b) is the greatly less-preferred reading in which *old* is a noun and *man* is a transitive verb. This is the reading that is necessary to parse (433), but it is locally overlooked in favor of (434a). Thus a garden-path effect results. This effect is explained by the lexical preferences associated with the words *old* and *man*. The word *old* is strongly preferred as an adjective and the word *man* is strongly preferred as a noun. As a result there is a very high load associated with (434b), which contains the unpreferred readings of each of these words. On the other hand, (434a) is associated with a much lower load, since it contains the preferred readings of these words. Because of the large load difference, only (434a) is maintained and a garden-path effect results.

Note that the garden-path effect disappears if the lexical preferences are not present:

(435) a. The desert trains young people to be especially tough.
b. The desert trains are especially tough on young people.

The word *trains* has noun and verb readings and the word *desert* can function adjectivally as in (435a) or nominally as in (435b). Since neither of these ambiguities is associated with a strong lexical preference, no garden-path effect is induced in either (435a) or (435b). See Section 6.9.2 for more complete explanations of the lack of garden-path effects in sentences with lexical ambiguities as in (435a) and (435b).

Furthermore, note that the preference to classify the word *old* as an adjective can be overcome when the word *man* is replaced with a word with the same categorical ambiguities but with the opposite preferences:[155]

(436) a. The old feed the young.
b. The old feed made the horse sick.

The input string *the old feed* has the same ambiguities as the string *the old man*, but there is no strong preference for either structure, as is evidenced by the lack of garden-path effect in either of the sentences in (436). While the word *old* is strongly preferred as an adjective, the word *feed* is also strongly preferred as a verb. Since each of the two possible structures for *the old feed* includes one preferred and one unpreferred reading the effects cancel each other and each can be parsed without garden-path effect.[156]

---

[155] The examples in (436) were pointed out to me by Robin Clark.

[156] For further information on lexical effects during parsing, see Clifton, Frazier & Connine (1984), Mitchell & Holmes (1985), and Mitchell (1989).

### 9.3. Semantic and Pragmatic Effects

Semantic and pragmatic effects can also over-ride what would be syntactic preferences. This section surveys a number of classes of examples whose syntactic structure is the same as earlier examples presented in this thesis, but whose preference is different, due to semantic and pragmatic factors. In order to account for the data, it is assumed that semantic/pragmatic markedness is associated with appropriate processing load. However, no particular load-bearing semantic or pragmatic properties are explicitly proposed; it is left to future work to discover these properties along with their corresponding loads.

#### 9.3.1. Lexical Ambiguity

Milne (1982) gives examples of lexical ambiguity which is resolved by appeal to knowledge of the outside world together with knowledge of the meanings of the words:[157]

(437) a. # The chestnut blocks fell off the table.
b. # The granite rocks during the earthquake.
c. # The aluminum screws into the opening.

Each of the sentences in (437) is ambiguous over the first three words between NP and (incomplete) sentential readings. The NP reading is locally preferred in (437b) and (437c), but it is the verbal reading that is required for a grammatical parse of each. The sentential reading is locally preferred in (437a), but here the NP reading is required. Hence each of these sentences induces a garden-path effect.

These effects cannot be explained by lexical preference. Consider (437c). Both adjectival and nominal lexical entries for *aluminum* are roughly equally weighted; that is, neither is rare as in the case of the nominal reading of *old*. Similarly, neither the nominal nor the verbal entries for *screw* is rare, so that these two entries are also roughly equally weighted. Thus, based solely on lexical preferences, neither the NP nor the VP reading of *the aluminum screws* should be preferred. However, the NP reading is preferred. This fact can be explained by appeal to semantic and pragmatic knowledge. Given some world knowledge about aluminum and screws, one knows that aluminum is a reasonable material from which to make screws. Thus the semantic/pragmatic structure representing *the aluminum screws* is associated with a small processing load. On the other hand, it is pragmatically odd for a mass noun phrase like *the aluminum* to be the subject of a verb like *screws* as either the AGENT, *instrument* or *theme*: the selectional restrictions of the verb *blocks* are violated (Katz & Fodor (1963), Katz & Postal (1964), Chomsky (1965)). As a result, there is a high load associated with this semantic/pragmatic structure, and the strong local preference for the NP reading of *aluminum screws* is explained.

Similarly, while a chestnut satisfies the selectional restrictions on the subject of the verb

[157] See Carlson & Tanenhaus (1988) for evidence in which the ambiguous lexical items are of the same category.

*blocks*, it is pragmatically strange for blocks to be constructed from chestnuts.[158] Thus there is a high semantic/pragmatic load associated with the NP reading of *the chestnut blocks*, while the sentential reading has a much lower cost. Hence the local preference in (437a) is explained.

### 9.3.2. Prenominal Modifier / Relative Clause Ambiguities

Consider (438) ((438b) adapted from Kurtzman (1985)), in which a syntactic preference is over-ridden by pragmatic factors:

(438) a. # The cotton clothing is made of grows in Mississippi.
b. The cigarette teenagers smoke most is Camel.

While there is a strong syntactic preference to treat *the cotton clothing* in (438a) as a single noun phrase (see Section 6.9.1), no such preference is present for the input string *the cigarette teenagers* in (438b). The explanation for this fact lies in semantics/pragmatics. There are a number of possible relationships between modifying and head nominals in noun-noun compounds (Downing (1977), Levi (1978), Finin (1980)). One such relationship is, loosely, that the referent of the head noun is composed of the referent of the modifying noun. This is the relationship between the two nouns in the NP *the cotton clothing* since clothing can be made of cotton. Thus the NP *cotton clothing* is associated with a low processing load with respect to semantics/pragmatics. On the other hand, the noun *cigarette* is a pragmatically marked modifier for *teenagers*: there is no relationship that is standardly shared between noun-noun compounds that these two nouns also share. In particular, teenagers are not normally composed of cigarettes. Thus there is a high processing load associated with the semantics/pragmatics of this NP. As a result, the syntactic preference for this structure is removed; the more highly marked syntactic reading, in which *teenagers* is the subject of a relative clause modifying phrase, can be maintained. Thus there is no garden-path effect in (438b).

### 9.3.3. Matrix Clause / Reduced Relative Clause Ambiguities

Kurtzman (1985) gives examples of sentences with a matrix clause / reduced relative clause ambiguity, in which the preferred resolution is determined by semantic and pragmatic properties. Consider the contrast between (439) and (440):

(439) a. The intelligent scientist examined with a magnifier the leaves from the largest tree he could find.
b. # The intelligent scientist examined with a magnifier was crazy.

[158] The word *chestnut* also has an adjectival lexical entry which indicates a color. I assume that this lexical entry is an unpreferred one, so that lexical preferences will account for its low priority.

(440) a. The intelligent alien examined with a magnifier the leaves from the largest tree he could find.
b. The intelligent alien examined with a magnifier was crazy.

In a grammaticality judgment task, Kurtzman (1985) found that while (439b) induces a garden-path effect, (440b) does not, despite its identical structure.[159] The only difference between the two sentences is the matrix subject of each: *scientist* in (439b) and *alien* in (440b). The difference in local preferences is given by semantic/pragmatic factors: while aliens are plausible themes of examination, scientists are not: scientists are most plausibly agents of investigation. Thus there is a relatively high processing load associated with the structure in which the scientist is the theme of examination, and the local preference is explained.

Related evidence is provided by Trueswell, Tanenhaus & Garnsey (submitted) who examined eye-movements in their experiments.[160] Trueswell *et al* looked at sentences like the following (originally from Ferreira & Clifton (1986)):

(441) a. The defendant examined by the lawyer turned out to be unreliable.
b. The evidence examined by the lawyer turned out to be unreliable.

While neither of the sentences in (441) causes a conscious garden-path effect, (441b) is processed more quickly than (441a). In (441a), the matrix NP *the defendant* is plausible as both AGENT and THEME of the verb *examined*, so that both matrix verb and reduced relative clause attachments are possible at this point. In (441b), on the other hand, the matrix NP *the evidence* is plausible only as THEME of the verb *examined*, since the selectional restrictions on the subject of the verb *examine* require an animate agent. The difference in the speed of parsing (441a) and (441b) is therefore explained by the semantic/pragmatic differences between the two local ambiguities: (441b) is parsed faster that (441a) since fewer structures are maintained in the parse of (441b). Only one representation is maintained once the word *examined* has been processed, whereas in (441a) a second representation is also maintained at the corresponding parse state.

### 9.3.4. Prepositional Phrase Attachment

While the Properties of Thematic Reception, Lexical Requirement and Recency Preference are important in determining the site of prepositional phrase attachment, semantic and pragmatic factors are also heavily involved (Schubert (1984, 1986), Kurtzman (1985), Wilks,

[159]See Crain & Steedman (1985) for similar results. See Section 6.5 for an explanation of the syntactic effects in sentence pairs like those in (439) and (440).

[160]Rayner, Carlson & Frazier (1983) and Ferreira & Clifton (1986) also observed eye-movements in studies of matrix / reduced relative clause ambiguities and found no effect of semantic and pragmatic information on the syntactic preference. The materials in both of these studies were severely flawed, however. See Kurtzman (1985) and Hirst (1987) for criticisms of the Rayner *et al* experiment and see Trueswell *et al* (submitted) for criticisms of the Ferreira & Clifton experiment.

Huang & Fass (1985), and Taraban & McClelland (1988, 1990), *cf.* Rayner, Carlson & Frazier (1983)).

#### 9.3.4.1 Locative PP Attachment

Consider the following examples (adapted from Wilks, Huang & Fass (1985)):

(442) a. John found the woman in the park.
b. John loved the woman in the park.

In (442a) the preferred attachment for the adjunct PP *in the park* is as modifier of the VP headed by *found* rather than as modifier of the NP *the woman.* This preference is the opposite of what would be expected if the Property of Recency Preference were the only factor involved in the determination of attachment sites. Note that when the eventive verb *found* is replaced by the stative verb *loved*, the preferred attachment for the PP *at the park* shifts to the NP *the woman.* A similar preference is present even if the event attachment site is a noun rather than a verb:

(443) a. the victory of the Americans in Europe
b. the shortage of food in Africa

Consider the noun phrase in (443a). There are two possible attachment sites for the PP *in Europe*: as modifier of the noun *Americans* or as modifier of the higher noun *victory.* Although the noun *Americans* occurs more recently in the input string, the attachment to the event nominal *victory* is preferred.

Note that when the ambiguity in attachment sites does not involve a variance in eventiveness, the preferred attachment is no longer to the matrix verb:

(444) a. I knew the woman in the kitchen.
b. John believed the secretary at the office.

For example, in (444a), the PP *in the kitchen* is preferred as modifier of the NP *the woman* over modifier of the VP headed by *knew.* The preference in (444b) is accounted for in a similar manner. In order to account for this data, I import the insight of Wilks *et al* into the framework presented here (*cf.* Abney (1989), Frazier (1990)). They propose that the preference for attachment of the locative PP *in the park* to the verb *found* in (442a) is given by the preference semantics of the lexical items in question. An eventive verb like *found* has a preference to be associated with locative information, while the noun *woman* does not. Translated into terms appropriate to the work presented here, a significant processing load is associated with the attachment of the locative modifier to the noun *woman* given the existence of a preferred attachment site, the eventive verb *found.* As a result of this preference, no additional processing load is associated with the attachment of the locative PP *in the park* to the VP headed by *found*, whereas a significant load is associated with the attachment of the same PP to the NP *woman.* If we assume that the load associated with this preference is greater than that associated with Recency Preference, then we have an explanation of the preferred interpretation in (442a).

#### 9.3.4.2 Direct Object Variation

A self-paced reading experiment conducted by Taraban & McClelland (1990) demonstrates the necessity to consider semantic and pragmatic factors when studying attachment preferences (see also Taraban & McClelland (1988), *cf.* Rayner, Carlson & Frazier (1983)). This experiment varied the direct object of the matrix VP as is demonstrated by the examples in (445):

(445) a. The dictator viewed the masses from the steps, but he was not very sympathetic.
b. The dictator viewed the masses from the city, but he was not very sympathetic.
c. The dictator viewed the petitions from the prisoners, but he was not very sympathetic.
d. The dictator viewed the petitions from the podium, but he was not very sympathetic.

The preferred attachment of the PP depended on the semantics/pragmatics of the matrix verb object. If the object of the verb *viewed* were *the masses*, then the preferred attachment of the following PP initiated by the words *from the* is to the verb. If, on the other hand, the object of the verb *viewed* were *the petitions*, then the preferred attachment of the following PP is to the object NP. These preferences were observed by measuring reading times of the PP object NP together with the words that followed it. Subjects were slower in reading these words in sentences like (445b) and (445d) than in (445a) and (445c), thus indicating that the matrix clause object affects the attachment preferences. In particular, the noun *petitions* seems to have a high preference for a source PP following it, while the noun *masses* does not.

Note that these results may be accounted for inside the framework presented here in much the same way as were the results from the previous section. I assume that there is a significant processing load associated with attaching an adjunct in a semantically unpreferred location when there is a more highly preferred attachment site available. Attachment of the PP headed by *from* to the verb *viewed* is preferred when the object NP is *masses* ((445a) and (445b)) since the verb *view* has a preference to be associated with a source, while the noun *masses* does not. Unlike *masses*, the noun *petitions* has a preference to associated with a source, so that both the matrix verb and direct object in (445c) and (445d) are preferred attachment sites in these sentences. In this case, the Property of Recency Preference takes over and the direct object attachment is preferred.

Related evidence is provided by the contrasts between each of the pairs of sentences in (446) and (447) ((446b) from Ford, Bresnan & Kaplan (1982); (446a) and (447) from Schubert (1986)):[161]

[161] Note that these data are only suggestive of local preferences since no on-line tasks were performed in these experiments.

(446) a. The women discussed the children in the kitchen.
b. The women discussed the dogs on the beach.

(447) a. John broke the vase in the kitchen.
b. John broke the vase in the corner.

While the preferred interpretation of (446a) links the PP *in the kitchen* to the verb *discussed*, this preference is greatly decreased in (446b) when the direct object is *the dogs* and the PP to be attached is *on the beach*. According to Ford, Bresnan & Kaplan (1982) the preference in this sentence links the PP *on the beach* to the NP *dogs* (*cf.* Schubert (1986) whose evidence indicates that there is no preference in (446b)). Similarly, the preference for attachment of the PP *in the kitchen* in (447a) is as modifier of the verb *broke*, whereas changing the PP object to *corner* alters the preference so that the attachment of choice is to the NP *the vase*. These preferences may be accounted for inside the framework given here in a manner similar to that of previous examples. In (446b), for example, the PP *on the beach* is at least as likely to be attached to the NP *the dogs* as to the matrix verb *discussed* because *dogs on the beach* make up a pragmatically plausible subject of discussion, more so than simply *dogs*. Thus there is a pragmatic load associated with attaching the PP *on the beach* to the matrix verb, and the more balanced attachment preference is explained (*cf.* Schubert (1984, 1986), Wilks *et al* (1985)).

### 9.4. Context and Discourse Effects

Based on his intuitions about how people process language, Winograd (1972) programmed his natural language understanding system SHRDLU so that context would be accounted for in order to disambiguate between syntactic parses. For example, consider (448):

(448) Put the block in the box on the table.

SHRDLU would look at the state of the current blocks world in order to see if there was already a block inside the box. If not, then the reading in which the prepositional phrase *in the box* modifies the NP *the block* is ruled out. While Winograd's intuition is plausible, he provides no psychological evidence in support of it. Thus no psycholinguistically relevant conclusions can be made based on Winograd's work.

Psychological evidence in support of the hypothesis that context has an immediate effect on attachment preferences is provided by Tyler & Marslen-Wilson (1977) and Marslen-Wilson & Tyler (1980, 1987). In one study these researchers showed how the content of the previous clause could change the preferred reading of an adjectival-gerundive ambiguity (Tyler & Marslen-Wilson (1977), Marslen-Wilson & Tyler (1987)). Example materials from their experiments are given in (449):

(449) a. If you walk too near the runway, landing planes...
b. If you've been trained as a pilot, landing planes...

After a sentence fragment similar to those in (449) was heard, a probe word appeared on the computer screen in front of the subject. The subject then read this word out loud. The experimenters measured the time from when the probe word first appeared on the screen until the subject named the word. The probe word was syntactically appropriate to complete either sentence, but was contextually more appropriate for one context rather than the other. The probe words for the partial sentences in (449) were *is* and *are*: the word *is* is an appropriate continuation if the word *landing* is categorized as a gerund, while the word *are* is an appropriate continuation if the word *landing* is categorized as an adjective, modifying the noun *planes*. It turned out that the probe word *is* was pronounced with a shorter delay following a sentence fragment like (449a) rather than following a fragment like (449b). Furthermore, the probe word *are* was pronounced more quickly following (449b) rather than (449a). The results of the experiments indicated that there is a contextual preference to treat the word string *landing planes* as a gerund following a context like (449a), while there is a similar preference to treat the word string *landing planes* as a noun phrase following a context as is given in (449b).[162]

Further evidence in support of the claim that context is involved in structural disambiguation is presented by Crain (1980), Crain & Steedman (1985), Altmann (1988) and Altmann & Steedman (1988).[163] Some of this evidence was presented in Section 2.6. Recall that sentences like (451a) and (451b) were placed in two different contexts: one following (450a) and the other following (450b):

(450) a. A psychologist was counselling a man and a woman. He was worried about one of them but not about the other.
b. A psychologist was counselling two women. He was worried about one of them but not about the other.

(451) a. The psychologist told the woman that he was having trouble with her husband.
b. The psychologist told the woman that he was having trouble with to leave her husband.

Both sentences in (451) are ambiguous up to the point of parsing the preposition *with*: the subordinate clause *that he was having trouble with* may attach as a complement of the verb *told*, or it may attach as a relative clause modifier of the NP *the woman*. (451a) is disambiguated in favor of the complement attachment while (451b) is disambiguated in favor of the relative clause modifier attachment. In the null context, (451a) is easy to process while (451b) causes a garden-path effect. However, following a context like (450b), in which there are two women in the discourse, the garden-path effect disappears for (451b) (*cf.* Kurtzman (1985)). In particular, Crain & Steedman found that sentences like (451b) were judged

[162]Townsend & Bever (1982) objected to the conclusions drawn by Tyler & Marslen-Wilson in the 1977 version of the experiment because of possible problems with some of the materials used. The experiment performed by Marslen-Wilson & Tyler (1987) attempted to answer these objections. See Marslen-Wilson & Tyler (1987) for a discussion of these issues.

[163]See Ferreira & Clifton (1986) and Clifton & Ferreira (1989) for arguments against the Crain, Steedman & Altmann position.

grammatical more often following their supporting contexts (sentence (450b)) than following non-supporting contexts (sentence (450a)). Similarly sentences like (451a) were judged grammatical more often following their supporting contexts (sentence (450a)) than following their non-supporting contexts (sentence (450b)). As a result, Crain & Steedman concluded that context was a crucial factor utilized by the human sentence processing mechanism.

In order to account for these results, I assume, following Crain & Steedman (1985), Altmann (1988) and Altmann & Steedman (1988) that there is a processing load associated with violating presuppositions associated with context. That is, given a context consisting of two women among other things, the noun phrase *the woman* by itself is not interpretable since there is more than one woman in the discourse. Thus there will be a high processing load associated with trying to interpret *the woman* as a complete NP inside a discourse containing more than one woman. As a result, the argument reading of *that he was having trouble with* in (451) will be associated with greater load in this context, and hence will be less preferred. The adjunct attachment will thereby become more preferred, and the encountered effects are explained.

# 10. Concluding Remarks and Future Work

## 10.1. Conclusions

This thesis has given a theory of sentence comprehension which attempts to explain a number of linguistic performance effects, including garden-path effects, preferred readings for ambiguous input and processing overload effects. The theory is based upon the assumption that there are processing costs associated with lexical, syntactic, semantic and pragmatic properties of linguistic structure. A global overload constant $K$ was proposed to account for processing overload effects while a preference constant $P$ was proposed to account for local preferences with respect to ambiguous input.

Four properties of syntactic structure were presented here. The first three properties – the Property of Thematic Reception (PTR), the Property of Lexical Requirement (PLR) and the Property of Thematic Transmission (PTT) – are based on the $\theta$-Criterion and Projection Principle from Government and Binding Theory. The fourth – the Property of Recency Preference (PRP) – prefers structures resulting from local attachments over those that result from more distant attachments. Inequalities in terms of these properties and the preference and overload constants were obtained by examining a wide range of data, predominantly from English. These inequalities are given in (452), where the loads associated with the PTR and PLR are combined into a single load, $x_{Int}$:

(452) a. $x_{Int} \leq P$
b. $2x_{Int} > P$
c. $|2x_{Int} - x_{TT}| \leq P$
d. $|x_{Int} - x_{TT}| \leq P$
e. $x_{Int} + x_{TT} \leq P$
f. $x_{RP} > P$
g. $x_{RP} - x_{Int} \leq P$
h. $5x_{Int} > K$
i. $4x_{Int} \leq K$
j. $4x_{Int} + x_{TT} > K$
k. $3x_{Int} + 2x_{TT} \leq K$

This set of inequalities is consistent. While depicting the solution space is not easy in two dimensions, we can still see that the set of inequalities is consistent by giving values for each of the properties that satisfy all the inequalities. One such solution is given in terms of the preference constant $P$:

(453) a. $x_{Int} = 0.6P$
b. $x_{TT} = 0.3P$
c. $x_{RP} = 1.2P$
d. $K = 2.5P$

For example, inequality (452k) is satisfied by the values in (453) since $3*0.6P+2*0.3P = 2.4P \leq 2.5P$.

To see more exactly how these values make the correct predictions with respect to the theory of processing breakdown, consider once again some of the sentences examined in the first chapter of this thesis. First let us look at sentence (454), which induces a garden-path effect (see Section 6.9.1):

(454) # The cotton clothing is made of grows in Mississippi.

At the point of processing the word *clothing* there are at least two interpretations for the input string, which are depicted in (455):

(455) a. $[_{IP}$ $[_{NP}$ the cotton clothing] ]
b. $[_{IP}$ $[_{NP}$ the $[_{N'}$ $[_{N'}$ cotton$_i$ ] $[_{CP}$ $[_{NP}$ $O_i$ ] $[_{IP}$ $[_{NP}$ clothing ] ]] ]]]

Structure (455a) is the single NP reading of *the cotton clothing*. The processing cost associated with this structure according to the properties presented here is $x_{Int}$ PLUs, since the NP *the cotton clothing* is in a thematic position but does not yet receive a thematic role. Thus the load associated with this structure is $0.6P$ PLUs assuming the property values given in (453).

Structure (455b) includes the NP *clothing* attached as subject of a relative clause modifier of the NP *the cotton*. The load associated with this structure is $3x_{Int}$ since it contains three NPs that require thematic roles but do not yet receive them. Thus the load associated with this structure is $1.8P$ PLUs according to the values in (453). The load difference between the two structures is therefore $1.2P$ PLUs, which is more than $P$ PLUs, the maximum maintainable load difference. Thus the more expensive structure, structure (455b), is pruned from the parse and a garden-path effect is correctly predicted.

In contrast, sentence (456) does not induce a garden-path effect because of the lack of similar local ambiguity:

(456) The cotton which clothing is made of grows in Mississippi.

The input string *the cotton which clothing* can only be construed in one way, and thus processing difficulty is avoided in (456).

Consider now the sentences in (457), each of which is locally ambiguous, but neither of which causes processing difficulty:

(457) a. The desert trains are especially tough on young people.
b. The desert trains young people to be especially tough.

The input string *the desert trains* is ambiguous between two readings: a noun-noun interpretation and a noun-verb interpretation. These two structures are given in (458):

(458) a. $[_{IP}$ $[_{NP}$ the desert trains] ]
b. $[_{IP}$ $[_{NP}$ the desert ] $[_{VP}$ trains $[_{NP}$ ] ] ]

Structure (458a) is associated with $x_{Int}$ PLUs since the NP *the desert trains* is in a position that can receive a thematic role, but it does not yet receive one. Structure (458b) is also associated with $x_{Int}$ PLUs since the verb *trains* has a currently unsatisfied lexical requirement. Thus the load difference between the two structures is 0 PLUs which is less than P PLUs and hence no processing difficulty is predicted with either structure, as desired.

Consider now (459), a sentence which causes processing difficulty in spite of the fact that it contains no local ambiguities:

(459) # The man that the woman that the dog bit saw likes fish.

The difficulty in processing this sentence is explained by the overload constant $K$: there exists a state in the parse of (459) such that the load associated with this state is larger than the processing overload constant $K$ PLUs. This state occurs after the second complementizer *that* has been input:

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

The load associated with this state is $5x_{Int}$ PLUs since 1) there are two lexical NPs, *the man* and *the woman*, which require thematic roles but do not locally receive such roles; 2) there are two non-lexical operators that require thematic roles but do not locally receive such roles; and 3) the second occurrence of the complementizer *that* has unsatisfied lexical requirements. Since $x_{Int} = 0.6P$ and $K = 2.5P$ by the solution space given in (453), the processing load associated with structure (460) is $3.0P$ PLUs, which is greater than $2.5P$ PLUs, the processing overload constant. Hence the processing overload effect in (459) is explained.

Similarly the maximal load associated with (461) is only $1.8P$ PLUs, and thus the lack of difficulty in processing this sentence is explained similarly:

(461) The man that the woman likes eats fish.

Since the system of inequalities obtained in this thesis is solvable, the theory of human sentence processing presented here is supported. Thus a theory of processing weight calculation, the preference factor $P$ and the overload factor $K$ provides a unified account of processing breakdown effects.

## 10.2. Future Work

While the theory of sentence processing given here is supported by a large quantity of data, much is still to be done. One area for future research is in determining what properties of linguistic structures beyond the syntactic ones discussed here are also associated with processing weight. While syntactic processing costs have been demonstrated here to be important in predicting processing breakdown, other linguistic properties are probably just as important, perhaps even more so. Aspects of this area of future research are examined in Chapter 9.

An additional area for future research is in testing the theory against languages other than English. While the theory of processing overload has been tested cross-linguistically to a small degree in Section 8.3, much additional work must be done, both in the area of processing overload and in the area of local preferences.

Furthermore, many questions are raised by the underlying assumptions of the sentence processing model given here. This thesis crucially assumes that there are processing weights associated with properties of linguistic structure. However, there are many ways other than the one proposed here in which a linguistic processing mechanism could utilize these weights to give a prospective theory of processing breakdown effects. I have made a number of simplifying assumptions here. First, I have assumed that calculating the total weight associated with a structure is a linear process: that is, the total weight associated with a structure is simply the sum of all the loads associated with properties of that structure. Furthermore, I have hypothesized the existence of two constants: a local preference constant $P$ and a processing overload constant $K$. Finally, I have assumed that the preference factor is an arithmetic preference factor rather than a geometric one.

While the assumption of linearity in load calculation has no obvious alternatives,[164] the same is not true of the other assumptions. For example, the assumption that the preference constant is an *arithmetic* constant has an obvious alternative: that this constant is *geometric* in nature. The choice of pursuing the arithmetic choice rather than the geometric one is based on practical issues. In investigating the arithmetic case, it is possible to discount properties that have an equal effect on two representations for the same input string when determining the load difference between the two structures, since these factors will cancel each other out when calculating a difference. However, it is not possible to discount these properties in the geometric case. If a property is associated with the same load with respect to two representations for the same input string, this property still must be considered since it will affect the overall ratio between the loads of the two structures. Thus all possible properties must always be considered when dealing with a geometric constant. Since it is probably the case that many properties are not yet known, it is therefore impractical to attempt explanations of preference effects in terms of a geometric constant.

The assumption that there are two separate constants, a preference constant $P$ and an overload constant $K$, as opposed to a single constant which explains both is one which stems from both practical and empirical considerations. That is, consider the hypothesis that there is no preference constant $P$, only an overload constant $K'$. In order to explain both processing overload phenomena and local preference phenomena, the overload constant $K'$ must apply across all representations. That is, let us assume the existence of an overload constant $K'$ which is the maximal quantity of PLUs associated with the total load of all representations that exist at one parse state. Furthermore let us assume that once the total load of all

[164]Of course, this is not to say that the linearity assumption is correct. It is only to say that I have as yet come across no reason to doubt it.

structures is greater than $K'$ PLUs, the most expensive structures are discarded or simplified in some way until the load is brought back down below $K'$ PLUs. For simplicity let us assume the Node Pruning Hypothesis: that the most expensive structures are discarded.[165] Thus a processing overload effect results if the load associated with a single structure is greater than $K'$. Furthermore, a garden-path effect results if the desired structure is pruned in favor of a locally more inexpensive structure. Thus the explanation of processing overload effects given in Chapter 8 would remain unchanged, but the explanations of preference effects given in Chapters 6 and 7 would be very different. This theory of processing overload and garden-path effects is preferable to the one given in this thesis on theoretical grounds since it contains fewer constants and is therefore simpler. However, this theory has a number of empirical flaws given the properties presented in this thesis. For example, consider the garden-path effects in (462):

(462) a. # The cotton clothing is made of grows in Mississippi.
b. # The furniture companies buy most is made by Sears.

Consider (462b). At the point of processing the word *companies*, there are at least two possible analyses, given in (463):

(463) a. $[_{IP}$ $[_{NP}$ the furniture companies ] ]
b. $[_{IP}$ $[_{NP}$ $[_{N'}$ $[_{N'}$ furniture$_i$ ] $[_{CP}$ $[_{NP}$ $O_i$ ] $[_{IP}$ $[_{NP}$ companies ] ]] ]]]

In structure (463a) *furniture companies* is analyzed as a single NP. In (463b) *companies* is analyzed as the subject of a relative clause modifier of the NP *furniture*. The load associated with (463a) is $x_{Int}$ PLUs, while the load associated with (463b) is $3x_{Int}$ PLUs. Thus the total load associated with both structures is $4x_{Int}$ PLUs. Hence the constant $K'$ must be less than or equal to $4x_{Int}$ PLUs in order to predict the garden-path effects in (462). However, it has been shown in Chapter 8 that $4x_{Int}$ PLUs is not enough to cause a processing overload effect. For example, $K'$ must be greater than $4x_{Int}$ PLUs in order to explain the lack of processing overload effects in sentences like those in (464):

(464) a. That the food that John ordered tasted good pleased him.
b. The possibility that the man who I hired is incompetent worries me.
c. What did the man that Mary likes eat?

There exist numerous other counter-examples like those in (462), so that collapsing the overload and preference constants into a single overload constant results in empirically undesirable consequences given the properties presented here.

However, the hypothesis that there is only one memory limitation constant should not yet be abandoned without further exploration. There are a number of ways in which this hypothesis can be salvaged. First of all, there may be as yet unconsidered structures in the processing of sentences like those in (462). Consider, for example, (462a). If there exists a structure associated with the input *cotton clothing* that is not already given in (463) and

[165] The Least Recent Nodes Hypothesis is also possible. See Chapter 8 for descriptions of these hypotheses.

has less load than is associated with structure (463b), then the lower bound on $K'$ increases. While I know of no such structure, one still might exist, in which case (462a) would no longer refute our hypothesis.

Secondly, the hypothesis can be maintained if the properties assumed in this thesis are either incomplete or incorrect as stated with respect to the supposed counter-examples. If there are relevant unconsidered properties associated with the processing of the sentences in (462), then the lower bound on $K'$ increases, and these examples no longer refute the hypothesis that there is only one memory limitation constant. Furthermore, if the properties that are presented here are not correct as stated then it is clear that the sentences in (462) need not be counter-examples to the $K'$ hypothesis. Thus it is not possible to refute the $K'$ hypothesis out of hand. Since the theory which contains only one memory limitation constant is simpler than the theory that contains two such constants, the single constant theory should not be abandoned without closer investigation. I leave this investigation to future work.

Whether or not it turns out that the existence of the preference constant $P$ is necessary, an additional locus of future work in this area concerns whether the processing overload constant $K$ applies across multiple representations. In all of the examples considered with respect to processing overload thus far I have only investigated solitary structures and their loads; I have not investigated whether the existence of the processing overload constant $K$ might also cause garden-path effects. That is, the preference constant $P$ and the overload constant $K$ might *both* cause local preferences, while remaining independent of one another. The preference constant $P$ would cause effects as demonstrated in this thesis, and the overload constant $K$ would cause additional preference effects when the total load across all representations becomes too great. For example, garden-path effects like those in (465) are currently unexplained in the current theory ((465a) from Marcus (1980)):

(465) a. # Have the soldiers given their medals by their sweethearts.
b. # Have the boys removed from the room.

The existence of garden-path effects in each of the sentences in (465) comes as a surprise to the approach presented here. While there are two possible structures for the initial substrings of the sentences in (465), the processing load differences for these structures never rises over $x_{Int}$ PLUs. Thus the garden-path effect in each is unexpected.

However, these effects might be explained by the application of the overload factor $K$ across all representations. Suppose that the properties of Thematic Reception and Lexical Requirement are restated so that loads are associated with lexical requirements and arguments assigned when there is more than one representation for an input string. Thus an argument is not associated with load via the Property of Thematic Reception if there is exactly one representation for the input string corresponding to that argument and the $\theta$-role assigner of the argument can be unambiguously determined. All other arguments are associated with load via the proposed PTR. Similarly, an element with lexical requirements is not associated with load via the Property of Lexical Requirement if there is exactly one representation for the input string corresponding to that element and its lexical requirement cannot be unambiguously determined. All other lexical requirement assigners are associated with load via the proposed PLR. Note that these proposed definitions do not alter the association of load with unassigned lexical requirements and arguments lacking thematic roles, so that basis for the explanation of the garden-path effects and processing overload effects given in this thesis remains.

As a result of the proposed definitions of the PTR and PLR, the garden-paths in (465) can be predicted by the processing overload constant $K$. For example, consider (465b) at the point of processing the verb *removed*. The two possible structures are given in (466):

(466) a. $[_{IP}$ $[_{VP}$ have $[_{IP}$ $[_{NP}$ the boys ] removed ]]]
b. $[_{CP}$ have $[_{IP}$ $[_{NP}$ the boys ] $[_{VP}$ removed ]]]

There are two different representations for the word *have*, one a matrix verb and the other an auxiliary. Each of these verbs has lexical requirements. Since there are two different representations for the word *have*, processing load is associated with each lexical requirement. There is also load associated with the matrix reading of the verb *have* since this reading is less common than the auxiliary form. Furthermore, the NP *the boys* is found in two representations, and is thus associated with load via the PTR. Finally, the verb *removed* is in two representations and thus its two thematic roles are associated with load via the PLR. Hence the total load across both representations is at least $5x_{Int}$ PLUs, enough for the overload constant $K$ to take effect. I assume that the load associated with structure (466a) is less than that associated with (466b) due mainly to the lexical preference of treating *have* as an auxiliary verb. Hence the garden-path effects in (465) have plausible explanations when the overload constant $K$ applies across all representations.

Clearly much more work needs to be done in this particular area to verify or falsify this explanation of the garden-path effects in (465). In particular, it is necessary to formalize the intuitive updated definitions of the PTR and PLR and, furthermore, to check that new definitions do not alter the desired predictions made by the definitions given in Chapter 5. I leave these issues for future research.

## 11. References

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

Abney, S.P, (1988) "On the Notions GB-Parser and Psychological Reality," in Abney, S.P. (ed.) "The MIT Parsing Volume, 1987-88," Parsing Project Working Papers, Number 1, Center for Cognitive Science, MIT, Cambridge, MA.

Abney, S.P. (1989) "A Computational Model of Human Parsing," *Journal of Psycholinguistic Research*, 18, 129-144.

Aho, A.V. & Ullman, J.D. (1972) *The Theory of Parsing, Translation, and Compiling, Volume 1: Parsing*, Prentice-Hall, Englewood Cliffs, NJ.

Altmann, G. (1988) "Ambiguity, Parsing Strategies and Computational Models," *Language and Cognitive Processes*, 3, 73-98.

Altmann, G. & Steedman, M. (1988) "Interaction with Context during Human Sentence Processing," *Cognition*, 30, 191-238.

Bach, E., Brown, C. & Marslen-Wilson, W. (1986) "Crossed and Nested Dependencies in German and Dutch: A Psycholinguistic Study," Language and Cognitive Processes, 1, 249-262.

Bar-Hillel, Y., Perles, M., & Shamir, E. (1961) *On Formal Properties of Simple Phrase Structure Grammars*, reprinted in *Readings in Mathematical Psychology.*

Belletti, A. & Rizzi, L. (1988) "Psych-Verbs and $\theta$-Theory," *Natural Language and Linguistic Theory*, 6, 291-352.

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

Berwick, R. (1987) "Principle-Based Parsing," MIT Artificial Intelligence Laboratory Technical Report 972, MIT, Cambridge, MA.

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

Blumenthal, A.L. (1966) "Observations with Self-Embedded Sentences," *Journal of Psychonomic Science*, 6, 453-454.

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.

Briscoe, E.J. (1983) “Determinism and its Implementation in PARSIFAL,” in K. Sparck Jones & Y. Wilks (eds.), *Automatic Natural Language Processing*, Ellis Horwood, West Sussex, U.K.

Carlson, G.N., & Tanenhaus, M.K. (1989) “Thematic Roles and Language Comprehension,” in W. Wilkins (ed.), *Syntax and Semantics Volume 21: Thematic Relations*, Academic Press, San Diego, CA.

Carpenter, P.A. & Daneman, M. (1981) “Lexical Access and Error Recovery in Reading: A Model Based on Eye Fixations,” *Journal of Verbal Learning and Verbal Behavior*, 20, 137-160.

Chomsky, N. (1956) “Three Models for the Description of Language,” *I.R.E. Transactions on Information Theory*, volume IT-2, 113-124, reprinted in R.D. Luce, R.R. Bush & E. Galanter (eds.) *Readings in Mathematical Psychology*, volume II, John Wiley, New York, NY.

Chomsky, N. (1957) *Syntactic Structures*, Mouton, The Hague, The Netherlands.

Chomsky, N. (1959a) “On Certain Formal Properties of Grammar,” *Information and Control*, 2, 137-167.

Chomsky, N. (1959b) “A Note on Phrase Structure Grammars,” *Information and Control*, 2, 393-395.

Chomsky, N. (1965) *Aspects of the Theory of Syntax*, MIT Press, Cambridge, MA.

Chomsky, N. (1970) “Remarks on Nominalization,” in R. Jacobs & P. Rosenbaum (eds.), *Readings in English Transformational Grammar*, Ginn, Waltham, MA.

Chomsky, N. (1973) “Conditions on Transformations,” in S. Anderson & P. Kiparsky (eds.), *A festschrift for Morris Halle*, Holt, Rinehart and Winston, New York, NY.

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

Chomsky, N. (1982) *Some Concepts and Consequences of the Theory of Government and Binding*, Linguistic Inquiry Monograph 6, MIT Press, Cambridge, MA.

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.

Chomsky, N. & Miller, G.A. (1963) “Introduction to the Formal Analysis of Natural Languages,” in R.D. Luce, R.R. Bush & E. Galanter (eds.) *Handbook of Mathematical Psychology*, volume 2, 269-321, John Wiley, New York, NY.

Church, K.W. (1980) *On Memory Limitations in Natural Language Processing*, MIT M.Sc. Thesis, Cambridge, MA.

Clark, R. & Gibson, E. (1988) "A Parallel Model of Adult Sentence Processing," *Proceedings of the Tenth Cognitive Science Conference*, Lawrence Erlbaum, Hillsdale, NJ.

Clifton, C., Frazier, L. & Connine, C. (1984) "Lexical and Syntactic Expectations in Sentence Comprehension," *Journal of Verbal Learning and Verbal Behavior*, 23, 696-708.

Cowper, E.A. (1976) *Constraints on Sentence Complexity: A Model for Syntactic Processing*, Ph.D. dissertation, Brown University, Providence, RI.

Crain, S. (1980) *Pragmatic Constraints on Sentence Comprehension*, Ph.D. dissertation, University of California, Irvine, CA.

Crain, S., & Fodor, J.D. (1985) "How Can Grammars Help Parsers?" in D. Dowty, L. Karttunen, & A. Zwicky (eds.), *Natural Language Processing: Psychological, Computational and Theoretical Perspectives*, Cambridge University Press, Cambridge, U.K.

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 Processing: Psychological, Computational and Theoretical Perspectives*, Cambridge University Press, Cambridge, U.K.

Cuetos, F., & Mitchell, D.C. (1988) "Cross-linguistic Differences in Parsing: Restrictions on the Use of the Late Closure Strategy in Spanish," *Cognition*, 30, 73-105.

Davidson, D. (1967) "Truth and Meaning," in Davidson, D. (1984) *Inquiries into Truth and Interpretation*, Clarendon Press, Oxford, U.K.

Downing, P. (1977) "On the Creation and Use of English Compound Nouns," *Language*, 53, 810-842.

Eady, J. & Fodor, J.D. (1981) "Is Center Embedding a Source of Processing Difficulty?," Presented at the Linguistic Society of America annual meeting.

Ferreira, F. & Clifton, C., Jr. (1986) "The Independence of Syntactic Processing," *Journal of Memory and Language* 25, 348-368.

Finin, T. (1980) "The Semantic Interpretation of Nominal Compounds," Report T-96, Coordinated Science Laboratory, University of Illinois, Urbana, IL.

Flickinger, D., Pollard, C. & Wasow, T. (1985) "Structure-Sharing in Lexical Representation," *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics.*

Fodor, J.A. (1983) *Modularity of Mind*, MIT Press, Cambridge, MA.

Fodor, J.A. & Garrett, M.F. (1967) "The Syntactic Determinants of Sentential Complexity," *Perception & Pychophysics*, 2, 289-296.

Fodor, J.D. & Frazier, L. (1980) "Is the Human Sentence Parsing Mechanism an ATN," *Cognition* 8, 417-459.

Fodor, J.D. (1985) "Deterministic Parsing and Subjacency," *Language and Cognitive Processes*, 1, 3-42.

Fodor, J.A., Bever, T.G. & Garrett, M.F. (1974) *The Psychology of Language*, McGraw-Hill, New York, NY.

Ford, M., J. Bresnan & R. Kaplan (1982) "A Competence-based Theory of Syntactic Closure," in J. Bresnan (ed.) *The Mental Representation of Grammatical Relations*, MIT Press, Cambridge, MA.

Forster, K.I. (1979) "Levels of Processing and the Structure of the Language Processor," in W.E. Cooper & E.C.T. Walker (eds.), *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*, Lawrence Erlbaum, Hillsdale, NJ.

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

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. (1990) "Parsing Modifiers: Special Purpose Routines in the Human Sentence Processing Mechanism?" in D.A. Balota, G.B. Flores d'Arcais & K. Rayner (eds.), *Comprehension Processes in Reading*, Lawrence Erlbaum, Hillsdale, NJ.

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

Frazier, L. & Rayner, K. (1982) "Making and Correcting Errors During Sentence Comprehension: Eye Movements in the Analysis of Structurally Ambiguous Sentences," *Cognitive Psychology* 14, 178-210.

Frazier, L. & Rayner, K. (1987) "Resolution of Syntactic Category Ambiguities: Eye Movements in Parsing Lexically Ambiguous Sentences," *Journal of Memory and Language* 26.

Freedman, S.A. & Forster, K.I. (1985) "The Psychological Status of Overgenerated Sentences," *Cognition*, 19, 101-131.

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

Garnham, A. (1985) *Psycholinguistics: Central Topics*, Methuen, London, U.K.

Garrod, S.C. & Sanford, A.J. (1985) "On the Real Time Character of Interpretation during Reading," *Language and Cognitive Processes*, 1, 43-59.

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

Gibson, E. (1990b) "Recency Preference and Garden-Path Effects" *Proceedings of the Twelfth Cognitive Science Conference*, Lawrence Erlbaum, Hillsdale, NJ.

Gibson, E. (1990c) "A Computational Theory of Processing Overload and Garden-Path Effects," *Proceedings of the Thirteenth International Conference on Computational Linguistics: COLING-90*, Helsinki University, Helsinki, Finland.

Gibson, E. (1991) "DiMorph: A Morphological Analyzer," Technical Report, Center for Machine Translation, Carnegie Mellon University, Pittsburgh, PA.

Gibson, E. (in preparation) "SPOT: A Subcategorization-driven Parser predicting Overload using Thematic information," Technical Report, Center for Machine Translation, Carnegie Mellon University, 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*, Ph.D. dissertation, University of Connecticut, Storrs, CT.

Grimshaw, J. (1981) "Form, Function and the language Acquisition Device," in C.L. Baker & J. McCarthy (eds.), *The Logical Problem of Language Acquisition*, MIT Press, Cambridge, MA.

Hirst, G. (1987) *Semantic Interpretation and the Resolution of Ambiguity*, Cambridge University Press, Cambridge, U.K.

Hobbs, J.R., Stickel, M, Martin, P. & Edwards, D. (1988) "Interpretation as Abduction," *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics*, Buffalo, NY.

Holmes, V.M., Kennedy, A., & Murray, W.S. (1987) "Syntactic Structure and the Garden Path," *Quarterly Journal of Experimental Psychology*, 39A, 277-294.

Huang, J. C.-T. (1982) *Logical Relations in Chinese and the Theory of Grammar*, Ph.D. dissertation, MIT, Cambridge, MA.

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

Jackendoff, R. (1987) "The Status of Thematic Relations in Linguistic Theory," *Linguistic Inquiry*, 18, 369-412.

Johnson-Laird, P. (1983) *Mental Models*, Cambridge University Press, Cambridge, U.K.

Joshi, A. (1990) "Processing Crossed and Nested Dependencies: An Automaton Perspective on the Psycholinguistic Results," *Language and Cognitive Processes*, 5, 1-28.

Kaplan, R.M. (1972) "Augmented Transition Networks as Psychological Models of Sentence Comprehension," *Artificial Intelligence*, 3, 77-100.

Kaplan, R.M. & Bresnan, J. (1982) "Lexical Functional Grammar: a Formal System for Grammatical Relations," in J. Bresnan (ed.) *The Mental Representation of Grammatical Relations*, MIT Press, Cambridge, MA.

Katz, J.J. (1964) "Semi-sentences," in J.A. Fodor & J.J. Katz (eds.), *The Structure of Language*, Prentice-Hall, Englewood Cliffs, NJ.

Katz, J.J. & Fodor, J.A. (1963) "The Structure of a Semantic Theory," *Language*, 39, 170-210.

Katz, J.J. & Postal, P. (1964) *An Integrated Theory of Linguistic Descriptions*, MIT Press, Cambridge, MA.

Kayne, R. S. (1984) *Connectedness and Binary Branching*, Foris, Dordrecht, The Netherlands.

Kennedy, A., Murray, W.S., Jennings, F., and Reid, C. (1989) "Parsing Complements: Comments on the Generality of the Principle of Minimal Attachment," *Language and Cognitive Processes*, 4, Special Issue, 51-76.

Kimball, J. (1973) "Seven Principles of Surface Structure Parsing in Natural Language," *Cognition* 2, 15-47.

Kimball, J. (1975) "Predictive Analysis and Over-the-Top Parsing," in J. Kimball (ed.), *Syntax and Semantics*, Vol. 4, Academic Press, New York, NY.

King, J. & Just, M.A. (submitted) "Individual Differences in Syntactic Processing: the Role of Working Memory," manuscript, Carnegie Mellon University, Pittsburgh, PA.

Knuth, D. (1965) "On the Translation of Languages from Left to Right," Information and Control, 8, 607-639.

Koster, J. (1978) "Why Subject Sentences Don't Exist," in S.J. Keyser (ed.), *Recent Transformational Studies in European Languages*, MIT Press, Cambridge, MA.

Kuno, S. & Oettinger, A.G. (1962) "Multiple Path Syntactic Analyzer," *Information Processing*, 62, 306-312.

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

Langendoen, T. (1975) , "Finite State Parsing of Phrase-Structure Language and the Status of Readjustment Rules in Grammar," *Linguistic Inquiry*, 6.

Larson, R.K. (1988) "On the Double Object Construction," *Linguistic Inquiry*, 19, 335-391.

Lasnik, H. & Kupin, J. (1977) "A Restrictive Theory of Transformational Grammar." *Theoretical Linguistics*, 4, 173-196.

Levi, J. (1978) *The Syntax and Semantics of Complex Nominals*, Academic Press, New York, NY.

MacDonald, M.C., Just, M.A. & Carpenter, P.A. (submitted) "Working Memory Constraints on the Processing of Syntactic Ambiguity," manuscript, Carnegie Mellon University, Pittsburgh, PA.

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

Marcus, M., Hindle, D. & Fleck, M. (1983) "Talking about Talking about Trees," in *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, Cambridge, MA.

Marslen-Wilson, W.D. & Tyler, L.K. (1980) "The Temporal Structure of Spoken Language Understanding," *Cognition*, 8, 1-71.

Marslen-Wilson, W.D. & Tyler, L.K. (1987) "Against Modularity," in J.L. Garfield (ed.), *Modularity in Knowledge Representation and Natural-Language Understanding*, MIT Press, Cambridge, MA.

McRoy, S.W. & Hirst, G. (1990) "Race-based Parsing and Syntactic Disambiguation," *Cognitive Science*, 14, 313-354.

Miller, G.A. (1956) "The Magical Number Seven Plus or Minus Two," *Psychological Review* 63, 81-96.

Miller, G.A. & Chomsky, N. (1963) "Finitary Models of Language Users," in R.D. Luce, R.R. Bush & E. Galanter (eds.) *Handbook of Mathematical Psychology*, volume 2, 419-491, John Wiley, New York, NY.

Miller, G.A. & Isard, S. (1963) "Some Perceptual Consequences of Linguistic Rules," *Journal of Verbal Learning and Verbal Behavior*, 2, 217-228.

Milne, R. (1982) "Predicting Garden Path Sentences," *Cognitive Science*, 6, 349-373.

Milne, R. (1986) "Resolving Lexical Ambiguity in a Deterministic Parser," *Computational Linguistics*, 12, 1-12.

Milsark, G.L. (1983) "On Length and Structure in Sentence Parsing," *Cognition*, 13, 129-134.

Mistler-Lachman, J.L. (1975) "Queer Sentences, Ambiguity and Levels of Processing," *Memory and Cognition*, 3, 395-400.

Mitchell, D.C. (1989) "Verb Guidance and Other Lexical Effects in Parsing," *Language and Cognitive Processes*, 4, Special Issue, 123-154.

Mitchell, D.C., Cuetos, F. & Zagar, D. (1990) "Reading in Different Languages: Is there a Universal Mechanism for Parsing Sentences?" in D.A. Balota, G.B. Flores d'Arcais & K. Rayner (eds.), *Comprehension Processes in Reading*, Lawrence Erlbaum, Hillsdale, NJ.

Mitchell, D.C. & Holmes, V.M. (1985) "The Role of Specific Information about the Verb to Parse Sentences with Local Structural Ambiguity," *Journal of Memory and Language*, 24, 542-559.

Moore, T.E. & Biederman, I. (1979) "Speeded Recognition of Ungrammaticality: Double Violations," *Cognition*, 7, 285-299.

Ni, W. & Crain, C. (1989) "How to Resolve Structural Ambiguities," *Proceedings of the Twentieth North East Linguistic Society Conference*, Pittsburgh, PA.

Norvig, P. (1988) "Multiple Simultaneous Interpretations of Ambiguous Sentences," *Proceedings of the Tenth Cognitive Science Conference*, Lawrence Erlbaum, Hillsdale, NJ.

Pereira, F.C.N. (1985) "A New Characterization of Attachment Preferences," in D. Dowty, L. Karttunen, & A. Zwicky (eds.), *Natural Language Processing: Psychological, Computational and Theoretical Perspectives*, Cambridge University Press, Cambridge, U.K.

Pesetsky, D, (1983) *Paths and Categories*, Ph.D. dissertation, MIT, Cambridge, MA.

Pollard, C. (1984) *Generalized Context-Free Grammars, Head Grammars and Natural Language*, Ph.D. dissertation, Stanford University, Stanford, CA.

Pollard, C. & Sag, I. (1987) *An Information-based Syntax and Semantics*, CSLI Lecture Notes Number 13, Menlo Park, CA.

Pollock, J.-Y. (1989) "Verb Movement, Universal Grammar and the Structure of IP," *Linguistic Inquiry*, 20, 365-424.

Pritchett, B. (1987) *Garden Path Phenomena and the Grammatical Basis of Language Processing*, Harvard University Ph.D. dissertation, Cambridge, MA.

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

Proudian, D. & Pollard, C. (1985) "Parsing Head-Driven Phrase Structure Grammar," *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics.*

Rayner, K., Carlson, M., & Frazier, L. (1983) "The Interaction of Syntax and Semantics during Sentence Processing: Eye Movements in the Analysis of Semantically Biased Sentences," *Journal of Verbal Learning and Verbal Behavior* 22, 358-374.

Rayner, K., & Frazier, L. (1987) "Parsing Temporarily Ambiguous Complements," *Quarterly Journal of Experimental Psychology*, 39A, 657-673.

Rothstein, S. (1983) *The Syntactic Forms of Predication*, Ph.D. dissertation, MIT, Cambridge, MA.

Schlesinger, I.M. (1968) *Sentence Structure and the Reading Process*, Mouton, The Hague, The Netherlands.

Schubert, L.K. (1984) "On Parsing Preferences," *Proceedings of Coling84*, Stanford, CA.

Schubert, L.K. (1986) "Are There Preference Tradeoffs in Attachment Decisions?" *Proceedings of the American Association of Artificial Intelligence 1986*, Philadelphia, PA.

Seidenberg, M.S., Waters, G.S., Sanders, M., & Langer, P. (1984) "Pre- and Post-Lexical Loci of Contextual Effects on Word Recognition," *Memory and Cognition*, 12, 315-328.

Shieber, S. (1983) "Sentence Disambiguation by a Shift-Reduce Parsing Technique," *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, Cambridge, MA.

Shieber, S. (1986) *An Introduction to Unification-based Approaches to Grammar*, CSLI Lecture Notes Number 4, Menlo Park, CA.

Shillcock, R. (1982) "The On-line Resolution of Pronominal Anaphora," *Language and Speech*, 25, 385-402.

Smith, N. (1986) review of Berwick & Weinberg (1984) *The Grammatical Basis for Linguistic Performance*, *Journal of Linguistics*, 22, 222-229.

Steedman, M. & Altmann, G. (1989) "Ambiguity in Context: A Reply," *Language and Cognitive Processes*, 4, Special Issue, 105-122.

Stolz, W.S. (1967) "A Study of the Ability to Decode Grammatically Novel Sentences," *Journal of Verbal Learning and Verbal Behavior*, 6, 867-873.

Stowell, T. (1981) *Origins of Phrase Structure*, Ph.D. dissertation, MIT, Cambridge, MA.

Swinney, D.A. (1979) "Lexical Access during Sentence Comprehension: (Re)consideration of Context Effects," *Journal of Verbal Learning and Verbal Behavior*, 18, 549-569.

Tanenhaus, M.K., Carlson, G.N. & Seidenberg, M.S. (1985) "Do Listeners Compute Linguistic Representations?" in D. Dowty, L. Karttunen, & A. Zwicky (eds.), *Natural Language Processing: Psychological, Computational and Theoretical Perspectives*, Cambridge University Press, Cambridge, U.K.

Tanenhaus, M.K., Leiman, J.M., & Seidenberg, M.S. (1979) "Evidence for Multiple Stages in the Processing of Ambiguous Words in Syntactic Contexts," *Journal of Verbal Learning and Verbal Behavior*, 18, 427-440.

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.

Taraban, R. & McClelland, J.R. (1990) "Parsing and Comprehension: a Multiple Constraint View," in D.A. Balota, G.B. Flores d'Arcais & K. Rayner (eds.), *Comprehension Processes in Reading*, Lawrence Erlbaum, Hillsdale, NJ.

Tomita, M. (1987) "An Efficient Augmented Context-Free Parsing Algorithm," *Computational Linguistics*, 13, 31-46.

Townsend, D.J. & Bever, T.G. (1982) "Natural Units Interact during Language Comprehension," *Journal of Verbal Learning and Verbal Behavior*, 28, 681-703.

Trueswell, J.C., Tanenhaus, M.K. & Garnsey, S.M. (submitted) "Semantic Influences on Parsing: Use of Thematic Role Information in Syntactic Disambiguation," manuscript, University of Rochester, Rochester, NY.

Tyler, L.K. & Marslen-Wilson, W.D. (1977) "The On-Line Effects of Semantic Context on Syntactic Processing," *Journal of Verbal Learning and Verbal Behavior*, 16, 683-692.

Waltz, D.L., & Pollack, J.B. (1985) "Massively Parallel Parsing: A Strongly Interactive Model of Natural Language Comprehension," *Cognitive Science*, 9, 51-74.

Wanner, E. (1980) "The ATN and the Sausage Machine: Which One is Baloney?" *Cognition*, 8, 209-225.

Wanner, E. (1987) "The Parser's Architecture," in F. Kessel (ed.) *The Development of Language and Language Researchers: Essays in Honour of Roger Brown*, Lawrence Erlbaum, London, U.K.

Weinberg, A.S. (1988) *Locality Principles in Syntax and in Parsing*, Ph.D. dissertation, MIT, Cambridge, MA.

Wilks, Y., Huang, X. & Fass, D. (1985) "Syntactic Preference and Right Attachment," *Proceedings of the International Joint Conference on Artificial Intelligence 1985*, Los Angeles, CA.

Winograd, T. (1972) *Understanding Natural Language*, Academic Press, New York, NY.

Yngve, V.H. (1960) "A Model and an Hypothesis for Language Structure," *Proceedings of the American Philosophical Society*, 104, 444-466.