PARSING AND INCREMENTALITY

by

David Andrew Schneider

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy with a major in Linguistics.

Summer 1999

PARSING AND INCREMENTALITY

by

David Andrew Schneider

Approved: ____________________

William J. Frawley, Ph.D.
Chair of the Department of Linguistics

Approved: ____

Conrado M. Gempesaw II, Ph.D.
Acting Vice Provost of Academic Programs and Planning

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____

Colin Phillips, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____

Robert Frank, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____

William J. Idsardi, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____

K. Vijayashanker, Ph.D.
Member of dissertation committee

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

There is a great deal of evidence that language comprehension occurs very rapidly. To account for this, it is widely, but not universally, assumed in the psycholinguistic literature that every word of a sentence is integrated into a syntactic representation of the sentence as soon as the word is encountered. This means that it is not possible to wait for subsequent words to provide information to guide a word s initial attachment into syntactic structure. In this dissertation I show how syntactic structures can be built on a word-by-word incremental basis. This work is motivated by the desire to explain how structure can be built incrementally.

A psycholinguistically plausible theory of parsing should generalize to all languages. In this work I show not only how head-initial languages like English can be parsed incrementally, but also how head-final languages like Japanese and German can be parsed incrementally. One aspect of incremental parsing that is particularly troublesome in head-final languages is that it is not always clear how a constituent should be structured in advance of the phrase-final heads. There is a significant amount of temporary ambiguity in head-final languages related to the fact that heads of constituents are not available until the end of the phrase. In this work I show that underspecification of the features of a head allows for incremental structuring of the input in head-final structures, while still retaining the temporary ambiguity that is so common in these languages. The featural underspecification allowed by this system is extended to categorial features; I do not assume that every head must always be specified for its category.

I assume that the incremental parser builds structures in accord with the principles of the grammar. In other words, there should be no need to submit a structure built by the parser to a separate grammar module to determine whether or not the sentence obeys the grammar. As one aspect of this, I show how wh-movement phenomena can be accommodated within the theory. As part of the treatment of wh-movement, constraints on wh-movement are incorporated into the system, thereby allowing the difference between grammatical and ungrammatical wh-movement to be captured in the parse tree.

In addition to being incremental and cross-linguistically generalizable, a parsing theory should account for the rest of human parsing behavior. I show that a number of the structurally-motivated parsing heuristics can be accommodated within the general parsing theory presented here. As part of the investigation of the incremental parser, experimental evidence is presented that establishes a preference for structure-preserving operations in the face of temporary ambiguity. In particular, the experiments show that once a commitment has been made to a particular analysis of a verbal argument, there is a preference to avoid reanalyzing the argument. This preference holds even though the reanalysis is not particularly difficult, and the analysis that is adopted in preference to the reanalysis disobeys a general parsing preference for attachments to recent material. Thus, it appears that existing structural assumptions are rejected only as a last resort.

Finally, to demonstrate the theory is explicit enough to make specific predictions, I implement portions of the theory as a computer program.

Chapter 1

INCREMENTALITY AND PARSING

1.1 Introduction and Motivation

One of the hallmarks of human language is that processing normally proceeds quickly and effortlessly. Generally, language is parsed and interpreted so easily that the process is not even noticed in everyday situations there are very few examples in which the level of difficulty rises to the point of being noticed consciously. The fact that language is processed so quickly and easily has led to a great deal of research to determine how sentences can be processed so easily.

This dissertation focuses on the question of how syntactic parsing takes place. There is little (if any) experimental work on human parsing that has shown substantial delays in language processing. Instead, parsing research consistently shows that language is processed very rapidly (Marslen-Wilson 1973, 1975, Stowe 1986, Trueswell, et al. 1993, Tanenhaus, et al. 1995, and many others). In contrast to computational models of parsing, many of which do not require that all words be incrementally integrated into a connected structure, psycholinguistic models generally assume that sentences can be interpreted at each stage of the parse (e.g. at every word). To allow for rapid interpretation, it has been argued that every word of a sentence must be integrated immediately into the syntactic representation of the sentence (Steedman 1989, Sturt and Crocker 1996). This means that integration of one word must take place before the next word is read in; it is not sufficient to simply buffer the words and process them at a later point on the basis of evidence elsewhere in the sentence. Examples (1) and (2) provide evidence that syntactic processing is carried out immediately.

(1) *Dorothy meet .the good witch.

(2) *The scarecrow believes that him needs a brain.

In (1), as soon as *meet* is processed (either heard or read), it is obvious to English speakers that there is a problem with subject-verb agreement. Likewise in (2), English speakers notice immediately that *him* is the wrong pronoun form for that syntactic position. Under certain common assumptions, it is not possible to determine whether or not a sentence is grammatical until the end of the sentence is reached (i.e. it is not possible to determine whether the valid prefix property holds). It might be argued that the determination of ungrammaticality could be made by some mechanism that does not reference linguistic structure (e.g. statistical properties of language). However, the sentence in (3) provides evidence that syntactic structure is available immediately to guide interpretation.

(3) The wizard$_i$ remembered that the tin man$_j$ talked to him$_{i/*j}$ about getting a heart.

One restriction on pronouns in English is that they cannot refer to NPs in the same clause, but are free to refer to any NP that is not part of the same clause as the pronoun. In (3) it is immediately obvious that *him* cannot refer to *the tin man* but can refer to *the wizard* (Nicol and Swinney 1989). This restriction on clausemate antecedents can only be taken advantage of if the syntactic structure is available to show which NPs are part of which clause.[1] Thus, the fact that co-reference is quickly restricted is a good indication that structure-building also takes place immediately.[2] I take the examples in (1)-(3) to

[1] A critic might argue that a pronoun simply cannot refer to the noun that immediately precedes it, or might argue that there is something about the words that intervene between *him* and *the tin man* that prevent co-reference. However, as shown in (i) below, neither of these arguments seems likely; Pronouns can refer to NPs immediately proceeding them given the proper syntactic conditions, even if the words between the pronoun and the preceding NP are exactly the same as in sentences where they may not corefer (as in (3)).

(i) The munchkin$_i$ who rescued the tin man$_j$ talked to him$_{*i/j}$ about getting a new heart.

[2] See Badecker and Straub (1999) and Straub and Badecker (1999) for evidence that all c-commanding NPs (even those grammatically prohibited from being antecedents), are at least initially considered as antecedents for pronouns and reflexives. These results are not in direct conflict with the claim that syntactic structure is immediately available, only with the claim that pronoun reference is determined *solely* on the basis of structure. Straub and Badecker find that the grammatically prohibited antecedents are correctly inhibited very quickly (<500 ms.), presumably on the basis of the syntactic structure.

show that the crucial word in each sentence (*meet, him,* and *him,* respectively) is integrated immediately into a syntactic structure containing the words previously processed.

This leads to one desideratum for a psycholinguistically plausible parsing theory: it should process sentences incrementally. All psycholinguistic models of parsing are incremental to one degree or another; in this dissertation I push incrementality to its logical limit by demonstrating how a parser can process every single word incrementally. Thus, no words will be left unattached at any stage of the parse. Before a new word can be integrated into the parse, all previous words must be integrated into a single structure (i.e. there cannot be two unconnected pieces of structure, each representing a different portion of the sentence).

Because the human capacity for sentence-processing is presumably invariant across languages (see Babyonyshev and Gibson 1995 for experimental evidence supporting one type of universality in the parser), it is also very desirable that a parsing theory be generalizable across languages. In Chapter 3 I focus on the difference between head-initial languages (e.g. English) and head-final languages like German (which is head-final in most embedded clauses) and Japanese (which is always head-final). As will be discussed below, some incremental parsers (Pritchett 1987, 1988, 1991, 1992, Stevenson 1994) are not able to parse head-final languages incrementally because they rely on information in the head of a phrase to structure the rest of the projection. For example, in Japanese, it is not at all unusual to encounter a sequence of three NPs before a verb is reached, as seen in (4).

(4) Bob-ga Mary-ni ringo-o ageta.
Bob-NOM Mary-DAT apple-ACC gave
Bob gave Mary the apple.

Chapter 3 presents arguments from the literature that sentences in head-final languages like Japanese are interpreted incrementally (e.g. in (4), the NPs are structured incrementally, well before the verb is reached). These arguments therefore favor an approach in which structure-building proceeds incrementally, even in head-final languages. One of the major goals of this thesis is to show how head-final languages can be processed just as incrementally as head-initial languages. The problem of not having information from a head to guide parsing does not generally arise in English because the head of any phrase (e.g. the verb) is available relatively early. For complements, the head of the phrase it will be part of has already been seen, so the information provided by the phrasal head can be used immediately; for specifiers, the information from the head is not available immediately (since the phrasal head has not been seen), but the information does become available fairly quickly, because the phrasal head is read in immediately after the specifier.[3] In the system presented here, every word is processed in a strictly incremental fashion. Thus, many items (specifiers (e.g. subjects) in all languages and complements (e.g. objects) in head-final languages) are integrated into syntactic representations before the head of the phrase containing them is reached.

A third desideratum for a parsing theory is that it account for human successes and failings. As is frequently pointed out, although the human sentence processor is in general quite quick and efficient, some sentences can cause conscious processing difficulty. Examples of these so-called (conscious) garden path sentences can be seen in (5)-(8) below (see Lewis 1993 for a wide-ranging review of garden path sentences).

(5) The horse raced past the barn fell. (Bever 1970)
(c.f. The car driven past the barn broke down.)

(6) While Mary was mending the sock fell off her lap. (Frazier 1978)
(c.f. While Mary was sleeping, the socks fell out of her lap.)

(7) The child put the candy on the table into his mouth. (Gibson 1991)
(c.f. The child put the candy from the store into his mouth.)

(8) Before the boy kills the man the dog bites strikes. (Warner and Glass 1987)
(c.f. Before the boy sleeps, the dog the man hits yelps.)

---

[3] If the specifier is very long, the problem becomes similar to the problem for complements in head-final languages the head that heads the phrase the specifier will be parsed into does not become available rapidly.

In each of these sentences, the globally incorrect analysis is pursued as a result of a temporary ambiguity, and the parser is unable to easily recover the correct analysis when the temporary ambiguity is disambiguated. The fact that these grammatical sentences cause parsing breakdown means that the mechanism responsible for quick, automatic parsing is not capable of dealing with all grammatical sentences. Since the automatic parser is not capable of dealing with all possibilities, it is clearly not an exhaustive parallel parser that pursues all possibilities to their logical conclusion. Although an exhaustive parallel parser might encounter difficulty (in the form of a long search), it should never fail completely. Thus, if the human parser were an exhaustive parallel parser, all sentences should be within the abilities of the parser.

Instead of full parallelism, two other types of models have been widely discussed in the literature: limited parallel models in which only some of the possibilities are pursued, and serial models in which only one parse is pursued at any given time.[4] Both of these approaches have been argued for in the literature, the parallel model by Gibson (1991 and subsequent), Pearlmutter and Mendelsohn (1999), MacDonald, et al. (1994), Gorrell (1987), Kurtzman (1985), and others. The serial approach has been argued for by Frazier (1978 and subsequent), Fodor and Inoue (1994, 1998), Gorrell (1995), Pritchett (1992), Weinberg (1993), Sturt and Crocker (1996, 1999), and others.

I assume a serial parser in this dissertation, but the ideas discussed are also applicable to limited parallel parsers. In particular, I show how syntactic structure can be built incrementally and completely bottom-up, and I also provide a mechanism for structure-building that retains (via underspecification of features in heads) some notion of the temporary ambiguities common in head-final structures. I assume that any change to the syntactic structure of a sentence has some processing cost, though different types of changes are associated with different processing costs. In the serial parser discussed

---

[4] It is unclear how network models like those of Tabor and Tanenhaus (in press), and Tabor, et al. (1997) fit into the serial/parallel distinction, since they don t clearly build syntactic structure at all. If anything, they are more like ranked parallel models than serial models, since the networks can encode many different interpretations at one time.

here, the featural underspecification has the effect of reducing the number of situations that require some sort of change (i.e. backtracking or repair) to the structure built incrementally. The ideas about how structure can be built incrementally should be applicable to both serial and parallel models, since the need to build structure incrementally is fundamental to both types of parsers. In the case of a parallel parser, the flexible structure produced by featural underspecification reduces the ambiguity load that the parser experiences during parsing.

Returning to the desideratum that the parsing model approximate human capabilities and inabilities, I show that the parser proposed here can account for many of the sentences that cause parsing breakdown, while simultaneously providing an account of sentences that can be processed without difficulty. I also show how this parser can incrementally build the structures necessary for a variety of syntactic constructions. Particular attention is paid to showing that the many wh-movement constraints in English and other languages can be captured in the system.

In summary, the goal of this work is to build a parser (called SPARSE[5]) that accounts for human syntactic processing in all of its richness. It should in general be quick, effortless, and incremental in both head-initial and head-final languages, but should still provide an account for difficulty observed in experimental investigations of various constructions. In an attempt to achieve these goals, considerable effort will be put into building structure that is compatible with a large number of alternative analyses (i.e. that retains some temporary ambiguity). While the goal is to try to minimize (or even eliminate) the structural changes/deletions that are required, complete elimination is not achieved; therefore methods to modify structure are also included. In order to demonstrate that the theory is adequate to parse both head-initial and head-final structures, parts of the theory were implemented in a computer program.

[5] For lack of a more original name, Schneider PARSEr.

1.2 Previous Work

This section begins with a quick review of some foundational work in psycholinguistic parsing, and then continues with discussion of the issue crucial to this work, how incremental analysis can be achieved. The section ends with a discussion of how parsing difficulty has been modeled through limits on structural change.

1.2.1 Parsing Heuristics

Early work on ambiguity resolution in parsing focused on heuristics for determining which analysis to pursue in temporarily ambiguous strings. One of the most well-known structure-based parsing strategies is *Minimal Attachment* of Frazier and Fodor (1978). Minimal Attachment states that the preferred interpretation for an incoming word is the one that adds the fewest number of additional nodes to the syntactic tree. Minimal Attachment effects were assumed to arise from the fact that it takes a uniform amount of time to build a new node, so an analysis requiring fewer nodes would be reached earlier than one requiring more nodes. This work assumed a parallel search for attachments, but serial transitions from word to word (i.e. once one parse was found for a given word, all other possible parses were ignored).

Minimal Attachment predicts that matrix verb attachments will be preferred to reduced relative clause analyses, as in (9), repeated from (5).

(9) The horse raced past the barn fell.

According to Minimal Attachment, the matrix clause reading of *raced* should be pursued, because it requires the addition of many fewer syntactic nodes than the reduced relative reading. Accordingly, difficulty is predicted when the verb *fell* is reached, because the analysis with a matrix verb reading for *raced* provides no spot for a sentence-final verb. Numerous other effects were argued to follow from the same principle.

Another parsing heuristic that appeared in the literature in the seventies is the strategy whereby new structure is attached at the lowest possible location in the tree.

This strategy was dubbed *Right Association* in Kimball (1973) and *Late Closure* in Frazier (1978). Although the accounts of Kimball and Frazier differ in the details and justifications, the basic idea is the same in both. Versions of this idea have been used in a number of other theories, under the names *Attach Low* (Abney 1989), *Recency* (Gibson 1991, Gibson, et al. 1996), and *Branch Right* (Phillips 1996).

This local attachment principle has been motivated by the fact that readers show a preference for attachments to more recent words in sentences in which attachments to both more and less recently encountered material are allowed by the syntax. Examples of this can be seen in (10) and (11) (cf. Kimball 1973).

(10) John said Bill left yesterday.

(11) Joe chewed the woman who wanted to take the cat out.

As the reader will notice, the adjunct *yesterday* in (10) is initially parsed as a modifier to the lower verb *left* rather than to the higher *said*. Likewise, the awkwardness noted in (11) at the particle *out* is due to the fact that the particle is normally attached to the verb *take*, rather than to the higher verb *chewed.*

Both of these parsing strategies (Minimal Attachment and Late Closure) have been widely cited in the parsing literature, but they do not comprise a comprehensive parsing theory. Instead, they are heuristics that can be used when the parser is faced with an ambiguity. There are many situations in which these heuristics do not provide proper accounts of the data (e.g. the argument attachment preferences discussed in /1.2.2.1 below). Crucially, the heuristics do not provide any mechanism for determining which attachments are possible. The problem of not providing a mechanism for determining possible attachments is a fairly general problem in the psycholinguistic literature many (but by no means all) of the psycholinguistic theories in the literature are concerned with resolving ambiguity once it has been identified, with little discussion of how the different structures involved in the ambiguity can be identified and constructed in the first place. Even among the psycholinguistic parsing theories that address how structure is built, many are not specific enough to actually explain the details of how structure is built. One of the goals of this work is to provide an account of incremental structure building that is explicit enough that it can be implemented in a computational model of parsing. For the most part, the theories of parsing that are specific enough to be implemented have been built with an eye towards efficiency, rather than with an eye towards modeling human parsing. Of course, there are a number of very explicit computational models that are designed to model human parsing. A number of these theories are discussed below.

1.2.2 Incrementality

This section provides an overview of a number of recent accounts of incremental parsing. They are reviewed with an eye towards the level of incrementality that they achieve, and a number of insights that will be expanded upon in this thesis are pointed out.

1.2.2.1 Pritchett s Principle-based Parsing

Pritchett (1987, 1988, 1991, 1992) proposes a strongly principle-based parsing model which operates solely by projecting phrasal structure strictly as determined by the lexical properties of heads and by licensing local attachments (Pritchett 1991, p. 252). The driving force behind parsing in this system is the need to satisfy as many grammatical constraints as possible at each stage of the parse (i.e. after each word is processed). To get an idea of how the system works, consider the example sentence in (12) below.

(12) Vampires were seen.

When *vampires* is processed, the head is projected up to an NP. Case-theory and the theta-criterion are left unsatisfied (as are any other features that might need to be licensed), since there is no head in the analysis that can satisfy these principles. Therefore, the grammatical constraints are satisfied to the maximum extent possible, and the parse proceeds to the next word.

When *were* is encountered, it is identified as an inflectional element and an IP (and the corresponding I^0) is projected. *Vampires* is licensed as IP specifier via case-assignment. This licensing relation must be formed because it allows a grammatical requirement to be satisfied. There is still no head to assign a theta-role to *vampires.* As a result, case theory, but not the theta-criterion, is satisfied.

Upon encountering *seen,* the parser recognizes it as a passive participle which can assign a theta role but cannot assign case. It projects a VP, which is licensed as the complement of *were*. Because *seen* assigns a theta-role, and the parser always attempts to maximally satisfy grammatical constraints, a chain is constructed between *vampires* and the object position of *seen*. The input then terminates, yielding a parse in which all grammatical principles are fully satisfied.

Pritchett (1992) shows that this system can account for parsing performance in a wide variety of structures. One fact that this system accounts for is the general preference to attach incoming constituents as arguments rather than adjuncts, even if the argument attachment is less local than the adjunct attachment. An example of this is found in (13), where the PP *with the binoculars* is preferably interpreted as an instrument of *saw* rather than as a modifier of the more recent NP *the man* (Frazier and Fodor 1978). This argument attachment is preferred because it allows for more theta-roles to be filled.[6]

(13) I saw the man with the binoculars.

Likewise in (14), an analysis in which *donations* is part of the NP object of *without* is preferred because it provides for more maximal satisfaction of grammatical constraints than an analysis in which *donations* is part of the matrix subject.

(14) Without her donations to the charity failed to appear.

When *donations* is encountered in the input, a analysis of it as part of the prepositional object allows it to have a theta role and case, while an analysis as matrix subject leaves it without a case-assigner or a theta role. Thus, it is parsed as part of the

---

[6] For Pritchett, maximal refers to the number of principles upheld/number of assignments that take place. Thus, a ditransitive use of a particular verb is preferred over a transitive use, because it satisfies more grammatical principles, thereby making its satisfaction maximal.

prepositional phrase, and difficulty is encountered later in the parse when there is no subject NP available for the verb *failed*.[7]

Pritchett s system accounts for additional facts which will not be discussed here; the reader is referred to the original source for details. Because parsing is based on the need to maximally satisfy grammatical principles, Pritchett s theory provides the basis for a theory of initial, incremental identification of parses, though he does not show how this principle could be formalized into an implementable parser.

Because of the requirement that parsing be based solely on the lexical properties of confirmed heads, the head-driven parsing model does not provide an incremental account of parsing in head-final languages. Pritchett (1991) states that in a sentence like (15), the three NPs that begin the sentence are buffered (i.e. remain detached from each other) until the first verb appears.

(15) Bill-ni Tom-ga Guy-o syookai suru to John-wa omotte-iru
Bill-DAT Tom-NOM Guy-ACC introduce IMPF COMP John-TOP think-ing
John thinks that Tom will introduce Guy to Bill.

Obviously, this is a significant deviation from full incrementality. Pritchett argues that there is no reason to assume that there is initially any structure to the three NPs at the beginning of (15), but a number of arguments that some syntactic structure is built in these sorts of NP sequences have appeared in the literature since 1992. Bader and Lasser (1994) provide experimental evidence from German that directly counters the claims of Pritchett s model. Bader and Lasser tested sentences like (16) and (17) in a word-by-word self-paced reading experiment.

(16) dass sie [nach dem Ergebnis zu fragen] tats chlich erlaubt hat
that she/her for the result to ask really allowed has
that she gave permission to ask about the result

[7] The fact that this preferred interpretation is difficult to revise is related to Pritchett s On-Line Locality Constraint, which limits the types of structural change/reanalysis that are possible. This will be discussed in ⁄ 1.2.3.3.

(17) dass [sie nach dem Ergebnis zu fragen] tats chlich erlaubt worden ist
that her/her for the result to ask really allowed been is
that permission has really been given to ask her about the result

In both of these sentences, the pronoun *sie* is ambiguous up until the auxiliary is reached between an object reading in the lower clause (as in (17)) and a subject reading in the higher clause (e.g. (16)). If Pritchett is correct, the pronoun should be left unprocessed until the first verb *fragen* is reached. When that verb (which was transitive-biased in the experiments) is reached, the principles that require maximal satisfaction of grammatical constraints should require that the pronoun be processed as part of the clause headed by *fragen* (as in (17)), since that will allow more grammatical constraints to be satisfied (i.e. it will allow the accusative case and the direct object theta-role of *fragen* to be assigned).

However, Bader and Lasser found that when the sentence is ultimately disambiguated (at the auxiliaries), the disambiguation in (17) that requires *sie* to be an object of *fragen* is read more slowly than the disambiguation in (16) that requires *sie* to be the subject of the higher clause. They also included unambiguous controls with unambiguously case-marked masculine pronouns in place of the ambiguous *sie*. In the unambiguous controls there was no difference between the two disambiguations.

Bader and Lasser interpret their results as evidence that *sie* is initially attached as a nominative subject, even before any verbs are reached.[8] Because it is already attached as a subject, it cannot be used as an object of *fragen*, even though the verb is biased towards having a direct object. Thus, Pritchett s claim that there is no structuring in advance of the heads, and that grammatical principles are maximally satisfied at all stages of parsing, is brought into question for these German sentences.

[8] As noted below in ⁄3.1, the assertion that *sie* is incrementally processed as a subject is open to other interpretations.

The need to incrementally interpret head-final structures is discussed further in ⁄ 3.1, which presents additional arguments about the need for incremental structuring of NP sequences in Japanese and German.

1.2.2.2 Stevenson s Competitive Attachment Model

Stevenson (1994) presents a hybrid connectionist architecture for parsing. In her system, nodes representing pieces of syntactic structure compete with one another for activation until such time as the network settles on a set of relations that constitute a single parse tree. Stevenson states that the method establishes syntactic relations both incrementally and efficiently (page iii). Because of the nature of competitive attachment, the system is able to model a variety of parsing effects without needing to posit explicit parsing strategies; all of the behaviors that are modeled elsewhere with heuristic parsing strategies (such as the Minimal Attachment and Late Closure heuristics discussed in ⁄ 1.2.1) are argued to fall out from the nature of the competition between elements. Stevenson s theory is more than just a theory of ambiguity resolution it also provides an explicit account of how structures are built and how possible attachment sites are identified.

In Stevenson s system, as in Pritchett s, no phrase can be built before the head of the phrase has been seen in the input. Stevenson limits the instantiation of nodes and phrases to those with overt evidence in the input. There are several reasons for this limitation. Among the reasons is the fact that it simplifies significantly the specification of the processing algorithm, since there is no need to differentiate between instantiated nodes and uninstantiated nodes. Other advantages include the fact that it reduces the number of nodes that need to be active in the network at any one time, and it eliminates the question of how to limit the amount of structure that can be hypothesized.

Although there are advantages to limiting nodes to only those unambiguously signaled in the input (by which Stevenson means that a category can only be instantiated if a word of that category has been encountered), it also has disadvantages. Like Pritchett s head-driven parser, Stevenson s parser is not able to achieve incrementality in head-final constructions (such as (15) above). Because all attachments must be based on features of a head (and such features/heads are not allowed to be intuited from other information), if the head is not present, no attachments can be made.

We will see later that by changing the definitions of overt evidence and syntactic head, Stevenson s idea of only instantiating nodes in the presence of overt evidence can be used profitably to allow incremental parsing of head-final constructions, while virtually eliminating the need to rescind predictions.

1.2.2.3 PARSIFAL

Marcus (1980) presents a parser (called PARSIFAL) that gives an explicit account of how structure can be built, in a system that also provides an account of parsing breakdown. To account for parsing breakdown (i.e. garden path sentences), Marcus assumes that once syntactic structure is built, it can never be retracted. Because no structure can ever be retracted in PARSIFAL, any mistakes that are made in the course of parsing lead to parsing breakdown. Marcus claims that PARSIFAL can account for a number of garden-path phenomena, precisely because it does sometimes make mistakes. In the vast majority of sentences, PARSIFAL does not make mistakes, just as the human parser rarely experiences conscious breakdown. In order to avoid making mistakes, PARSIFAL uses a three constituent buffer. The items in the three constituent buffer can be combined as needed, and can also be used to guide parsing of elements that precede them in the sentence. PARSIFAL uses a relatively unorthodox system of packets of pattern-action rules rather than the more common system of phrase structure rules and parsing heuristics. However, this difference in how structure is built is unimportant to the discussion at hand.

The fact that PARSIFAL uses a three-constituent buffer, however, is quite relevant to the present discussion. The three-constituent buffer means that very frequently incrementality is not observed. Because the buffer is three *constituents* rather than three *words*, the buffer can contain a large number of words. Marcus presents the following pair of readily understandable sentences to show how the buffer is used:

(18) Have the students who missed the exam take it today.

(19) Have the students who missed the exam taken it today?

These two sentences are identical for the first seven words, but have a different syntactic structure starting with the very first word. This means that the first seven words must be buffered (as parts of three constituents) before the very first word can finally be assigned its final analysis. This is a significant deviation from incrementality.

While Marcus abandoned the specific details of PARSIFAL within just a few years, the idea of structural determinism (that no structure can be retracted) formed the basis of description-theory (Marcus, Hindle, and Fleck 1983). that will be discussed below in /1.2.3.2 , /3.3 and /5.4.1.2 .

1.2.2.4 Categorial Grammar

Within the framework of Combinatory Categorial Grammar (CCG), Steedman (1993, 1996, to appear) develops a theory in which various apparently contradictory constituency facts can be easily accounted for. CCG is able to account for these different facts by virtue of a syntactic system that allows multiple derivations for each sentence, some right-branching and some left-branching. The basic idea is that words are associated with one or more types/categories, and the rules of the system frequently allow a given set of words to be combined in several different orders. CCG does not actually build phrase structure in the normal sense. Once a constituent is built, its internal structure is discarded. However, constituency relations can be determined by examining the derivation of the sentence, which shows the constituents that were built in the course of deriving the sentence.

For example, the sentence in (20) can receive both of the derivations shown below it.

(20) Dorothy saw him.

```
Dorothy       saw       him                        Dorothy   saw              him
 --------    ------------   -----                   ----------  -----             -----
   NP      (S\NP)/NP   NP                            NP      (S\NP)/NP      NP
           ------------------->                     ------->T
                  S\NP                             S/(S\NP)
   <---------------------                           --------------------------FC
              S                                                S/NP
                                                               ----------------------->
                                                                         S
```

In reading these derivations, one should know that each of the categories is a syntactic (and semantic) type that defines its combinatory possibilities. If the type is a simple argument, a primitive symbol (e.g. NP) is used to represent it. If the type is a function, it specifies what it combines with. A forward slash / specifies that the type combines with a category on the right, and a backslash \ means that the type combines with something on the left. Thus, the verb *saw*, of type (S\NP)/NP, first combines with an NP on its right.[9] This combination results in a constituent of type S\NP, which is a function that will return an S when it is combined with an NP on its left. If these were the only methods of combination in CCG, incrementality could not be straightforwardly achieved, because of the fact that the grammatical type specifies that the verb must first combine with its object before it can combine with the subject. Thus, the entire VP would need to be completed before the subject could be attached.[10]

There is an additional type-raising rule that allows an argument to specify what sort of function it will combine with. Type-raising is a device that can be applied to arguments to turn them into functions. Essentially, this amounts to specifying which categories a type-raised element can combine with. For example, the NP *Dorothy* can be type-raised to S/(S\NP), which means that it will combine with a VP (=S\NP) on its right to form an S. Through a composition operation, this rightward-looking NP can combine

[9] In CCG generally, elements combine first with the elements outermost in the type specification, rather than with the innermost types. E.g. an element of type X/(Y/Z) would combine with an element of type Y/Z to yield an element of type X.

[10] This would be similar to a na ve system using phrase-structure rules, in which a subject couldn t be combined with a verb until the entire VP has been completed.

with the verb, retaining any needs that the type of the verb might encode. In this way, the subject *Dorothy* can combine with the verb *saw* in (the right-hand derivation in) (20), to form a constituent that will be an S when it combined with an (object) NP.

Within CCG, it is not necessary to store hierarchical structure for interpretation, because the semantic information normally gleaned from a syntactic tree can be built during the derivation (see Steedman 1996 for details). Thus, the derivations shown here are just that histories of the derivation rather than constituent structures built up by the parser. However, the derivation does show which elements were constituents at some point in the derivation, and in that sense they correspond to standard constituency trees. To simplify the discussion that follows, I will use standard tree structures to show the history of a CCG derivation.

As can be seen above, there are two possible derivations for (20), one with the verb and object forming a constituent, the other with the subject and the verb forming a constituent. These different constituency possibilities allow for a very simple explanation of the constituency facts in (21) and (22). The sentence in (21) is an example of standard VP coordination, and (22) shows that non-standard subject-verb constituents can also be coordinated.

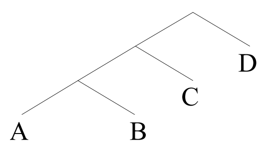
(21) Dorothy [liked the good witch] and [hated the wicked witch].

(22) [Dorothy knew] and [the lion recognized] the good witch.

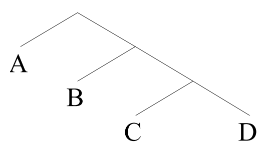
As in most other syntactic theories, Steedman assumes constituents can be coordinated whenever their categories are the same. Thus, by allowing the subject and verb to form a constituent, they can also be coordinated, as in (22). In that sentence, both *Dorothy demanded* and *the lion requested* are constituents of category S/NP (i.e. when an NP is seen on the right, it will combine with the S/NP to form an S.) By positing that the subject and verb make up a constituent of type S/NP, Steedman can allow subject-verb combinations to be coordinated, yet still retain the requirement for a direct object.

The addition of type-raising allows left-branching structures for English. This left-branching structure allows Steedman to easily account for a large amount of incrementality without needing to build partial constituents, since entire constituents can be built up incrementally in English.[11] This contrasts with parsers that incrementally build right-branching structures. Any parser that builds right-branching structures incrementally must build partial constituents in order to allow full incrementality (this parser included).

(23)



(24)



As can be seen in (23), in a left branching structure of the sort that might be built in a CCG, the strings A , A B , A B C , and A B C D are each constituents. In contrast, in the right branching structure (24) that would be built in most other grammar formalisms for the string A B C D , the strings A B and A B C are not constituents. Thus, a parser that is limited to building full constituents will not be able to incrementally process these right-branching structures. The only way to incrementally build the right-

---

[11] Type-raising is limited to arguments in order to avoid over-generation and possible infinite recursion. Because of this, full incrementality cannot be achieved, because categories like determiners cannot type-raise. If this restriction on type-raising is lifted, full incrementality can be achieved, but at the expense of allowing too much coordination. Milward (1994) proposes such an unrestrained system. Milward works in dependency grammar, but the same effects could be derived in a CCG if unlimited type-raising were allowed in a CCG. . Milward states that the account predicts that any substring of a sentence can coordinate with a parallel substring. The theory thus predicts that the following are all grammatical:

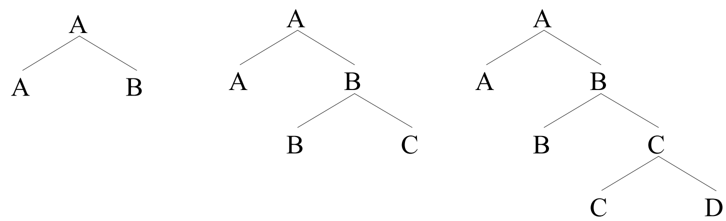
(ii) * Every [[boy believes] and [girl doubts]] scary ghost stories

(iii) *The children [[ate all of her good] and [discarded some of my burnt]] cookies.

(iv) *I saw [[a friend of]$_i$] and [a relative of]$_i$] Mary$_i$ s handbags.

branching structure in (24) would be to build partial trees that involve constituents that will not be present in the final structure. For example, the structures shown in (25) might be built in an incremental parse of the string A B C D . Notice that in this sort of structure, A B is temporarily a constituent, as is A B C .

(25)

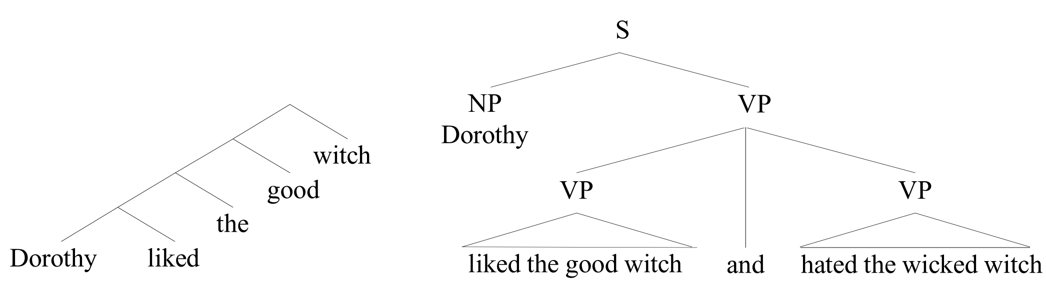


Although Steedman s structures do allow for significant incremental processing, and also allow substantial flexibility in constituency that is in accord with a wide range of constituency facts, the left-branching structure that is needed for incrementality in Steedman s approach precludes coordination for many constituents that traditional right-branching structures can accommodate quite easily, as seen in (26) below.

(26) Dorothy [liked the good witch] and [hated the wicked witch].

A completely left-branching structure, similar to what would be needed to parse this sentence incrementally using Steedman s approach, is shown on the left in (27). As can be seen, the VP *liked the good witch* is not a constituent, thus it cannot be coordinated with the second conjunct of the VP in (26).

(27)



On the right in (27) is the structure that would be required by Steedman to allow for the VP coordination. Notice that in this structure, the conjoined VP is not completed until both of the VP conjuncts have been built completely; therefore the entire VP structure cannot combine with the subject *Dorothy* until then. Obviously, this is a significant deviation from incrementality. The sentence in (28) provides evidence that sentences like (26) must be parsed incrementally. In particular, the fact that the reflexive *herself* in (28) can be interpreted immediately provides evidence that interpretation does not wait until the end of the conjoined VP.

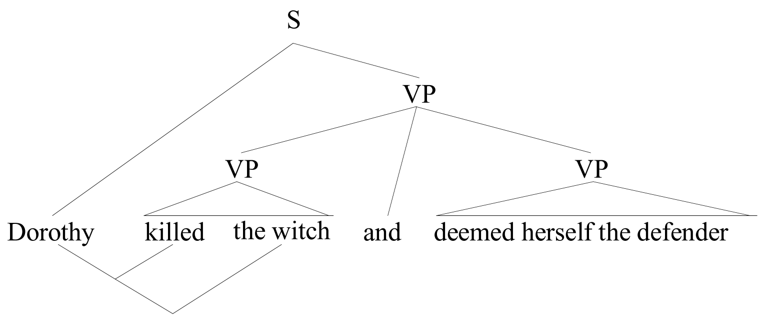
(28) Dorothy [[whispered to herself] and [pleaded to the wizard]].

One possible way around this non-incrementality is presented in Pareschi and Steedman (1987). They present a method that differs from the most straightforward application of CCG to parsing in that it allows incremental interpretation of the reflexive in (28). The theory developed by Pareschi and Steedman allows for a normal incremental parse of sentences like (28) up to the conjunction (along the lines of the left-branching structure in (27)), followed by the reconstruction of a VP for the first conjunct that allows for coordination of the two VPs. The specific mechanism is not discussed here; the reader is referred to the original paper for details.

To see how this would work, consider the VP coordination in (29), and the accompanying derivation tree in (30).

(29) Dorothy killed the witch and deemed herself the defender of the little people.

(30)



As can be seen from the tree below the words, the first conjunct can be parsed incrementally. The tree above the words shows that a VP can be constructed for *killed the witch* from the features of the lower tree for *Dorothy killed the witch*. This VP node can then be coordinated with the VP for the second conjunct, once the entire second VP has been built. However, this level of incrementality is not sufficient to explain the fact that *herself* can be interpreted immediately. In CCG, the interpretation of a reflexive relies on the fact that the reflexive and its antecedent are in a single syntactic structure. Thus, it should be impossible to correctly deduce that *herself* refers to *Dorothy* and not to *the witch* until the second VP conjunct is completed. However, it is clear as soon as *herself* is reached that *Dorothy* must be its antecedent.

Thus, even the additional flexibility that is allowed by the theory of Pareschi and Steedman (1987) is not enough to account for the incrementality facts presented above. Because of the fact that coordination cannot take place until constituents have been completed, the second conjunct of a coordinate structure cannot be interpreted incrementally.

1.2.2.5 Phillips s Left-to-Right Syntax

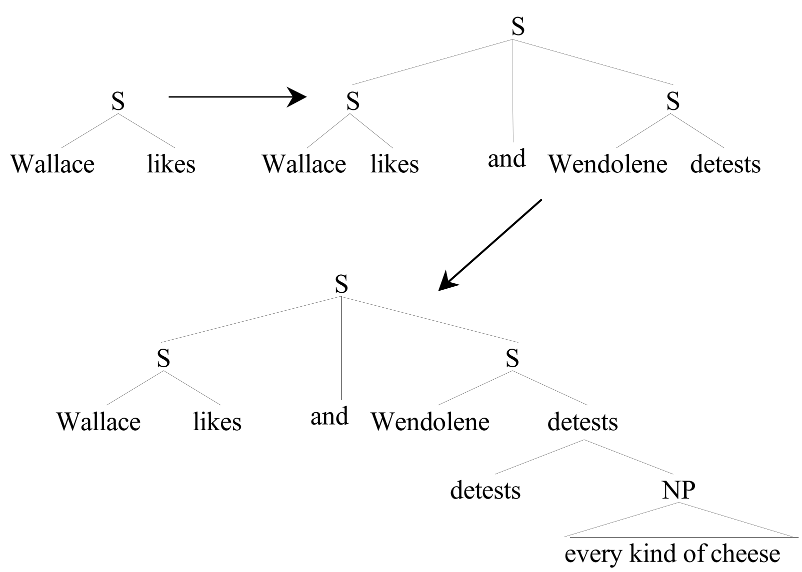
To account for the apparently contradictory constituency facts that CCG successfully handles, I rely upon the system developed in Phillips (1996, to appear). Phillips argues for an incremental (left-to-right) syntax. By using the sorts of structures shown in (25) as temporary, intermediate parse states, Phillips allows much more flexible constituency than normally assumed while maintaining a traditional right-branching structure for English. In the approach taken by Phillips, in contrast to that taken by Steedman, there is only one possible syntactic structure for any given sentence. Thus, Phillips does not rely on the fact that there are different ways to combine structures to achieve the structures needed to handle non-constituent constituency facts.

To account for non-constituent coordinations like (31), Phillips allows temporary constituents to coordinate.

(31) [[Wallace likes] and [Wendolene detests]] every kind of cheese.

Phillips builds bare phrase structures of the sort assumed in this thesis. The relevant steps in the incremental building of (31) are shown below in (32).

(32)



As can be seen in the trees in (32), the temporary constituent *Wallace likes* is coordinated with another identical constituent, *Wendolene detests*. Under a traditional bottom-up theory of structure-building, the words *Wallace likes* never form a constituent. However, because Phillips assumes that structure is built left-to-right, this sort of constituent is routinely available.

While Phillips does provide a principled account of the apparently contradictory constituency facts that motivate Steedman s CCG account, Phillips s account does not answer all of the questions posed in this work. Although Phillips does provide an principle to decide which parse will be pursued if multiple parses have been identified, he does not discuss how parses are initially identified. The principle used for ambiguity resolution, called BRANCH RIGHT, states that the attachment that uses the shortest path from the most recently processed word is preferred. As the reader can imagine, this principle favors local attachments over non-local attachments.

While it is clear what sort of structures Phillips would need to posit to process English incrementally, Phillips does not address the general question of how head-final languages can be accounted for incrementally. It is unclear how structure might be built incrementally for sentences in which a number of verbal arguments are received before the verb, such as (4) above. If these sentences are structured incrementally, as Phillips assumes, then some sort of structure must be built to allow the arguments to be interpreted incrementally. Unfortunately, Phillips does not address what that structure might be, or how it might be identified and built.

1.2.2.6 Left Attachment Parsing

In another system that is designed to achieve incremental structure building and interpretation, Stabler (1994) presents a method of identifying and building structures that is in many respects very similar to the one proposed in this dissertation. In the left-attachment parser Stabler modifies the left-corner parsing algorithm (Rosenkrantz and Lewis II 1970) to allow a more incremental parsing process than the standard left-corner algorithm provides. Left-corner parsers generally allow for more incremental parsing than typical bottom-up parsers by allowing some constituents to be attached into the main tree before the constituent has been completed.[12] The basic intuition behind a left-corner

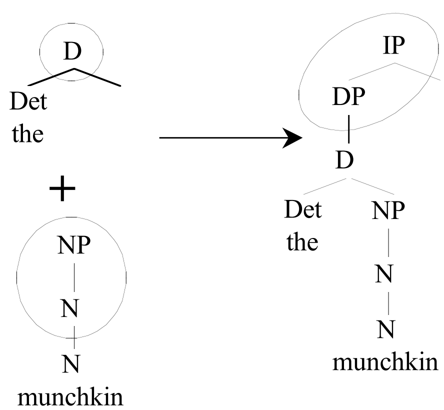
[12] Also psycholinguistically relevant, a number of researchers have also investigated the extent to which memory limitations and the structures required by left-corner parsers can provide an account for the difficulty of center-embeddings (Johnson-Laird 1983, Abney and Johnson 1991, Resnik 1992).

parser is the following: structure is built bottom-up, but after any constituent (the left corner) is completed, its parent node is built and potential siblings of the left corner are predicted.

(33) D _ Det NP
DP _ D
N _ N
NP _ N
IP _ DP I
Det _ *the*
N _ *munchkin*

For example, assuming the rules in (33), the word *the* allows a Det node to be built. Because Det is the left corner of the D rule, a D is also built. *Munchkin* causes the building of an NP, which can be attached as the right-hand member of a D . When the D is completed, a DP is built. Finally, because DP is the left corner of an IP, and IP node is also built. This process in seen in (34), where the circled material indicates nodes predicted at each step in the derivation.[13]

(34)



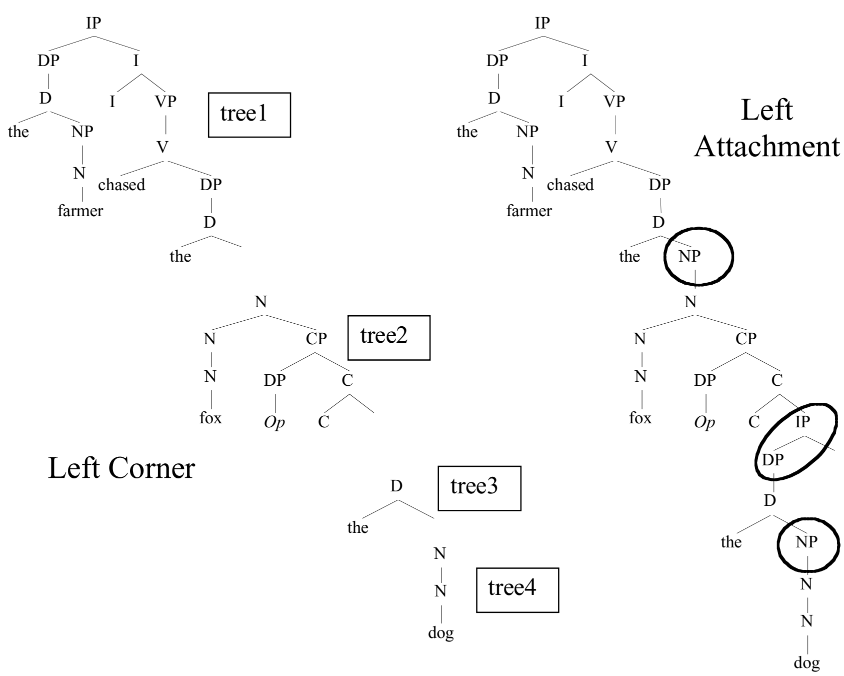
To see the limits of left-corner parsing, consider the sentence in (35), from Stabler (1994).

[13] In some versions of left-corner parsers, the sister nodes of the left corner (e.g. the NP sister to Det) would be predicted with their categories, while in other versions the categories themselves are not predicted. The arguments presented here apply regardless of whether or not the sisters are predicted.

(35) The farmer chased the fox the dog bit.

Consider how this sentence would be parsed by left-corner parser using a grammar that implements full X -templates. The left-hand tree in (36) below shows the parse produced before the final verb *bit*. Notice that there are four separate trees at this point that have not been connected. The uppermost tree, tree1, cannot be connected to tree2 because there is no NP node. No NP can be built up from the N in tree2 because the N is not yet complete.[14] Tree2 and tree3 cannot be connected because the DP and IP nodes are missing. They cannot be posited because the D has not been completed. Likewise, because the N in tree4 could potentially have a sister, it cannot be completed. As a result, tree3 and tree4 cannot be connected.

(36)



[14] Even if an NP were predicted left-corner, there would be no way to connect the NP and N , since the N has not been completed and the NP does not force any predictions below it, since it is a predicted element.

By making slight modifications to the left-corner algorithm, Stabler is able to produce the structure on the right, which consists of just one tree that contains all of the input so far.

The differences between Stabler s left-attachment parser and a more traditional left-corner parser are the following:

(a) if XP is a predicted node and if the next word is a head X, the X can be attached (through the X node that is predicted left-corner from the X) to the predicted node XP (predicting siblings at each level, left-corner-style)

(b) if XP is a predicted node and there is a rule XP_YP X and if the next word is a head Y, then Y can be projected to YP and attached as specifier of XP (predicting siblings at each level).

These two additions in the left-attachment parser allow it to posit the extra structure (circled in the tree on the right) needed to parse the sentence incrementally. Essentially, these modifications allow the parser to project an X^0 up further if that will allow the XP to be attached into the main tree. The modification made in step (a) merely allows a confirmed X^0 to project up to a maximal projection if that maximal projection is already predicted (i.e. it relaxes the condition on completions). An example of this is seen in (36), where an NP (circled) is built to connect tree1 and tree2. The task of connecting an X^0 to an XP is a non-issue in more recent versions of phrase structure (i.e. bare phrase structure in Chomsky 1995a), where there is no X^0/XP distinction. For this reason, I will not dwell on this issue here. How bare phrase structure is used in this work will be discussed below in ⁄2.3

The work done by (b) is somewhat more interesting. It accounts for the DP and IP nodes that are posited above *the dog* (between tree2 and tree3) in (36). This additional step essentially allows the IP to be built if the head above calls for an IP and a possible specifier for the IP can be predicted. As we will see below, I implement an idea similar to this extra material can be projected up from the new word in order to allow it to attach to the existing tree. The main difference between the way this is implemented in the two systems is that in the LA parser, structure is built both top-down and bottom-up, while in SPARSE, all new structure is projected bottom-up.

While the Left-Attachment parser does allow for substantial incrementality, it does not solve all the problems that SPARSE is designed to solve. One goal of this parser is to parse, where justified by human behavior, with minimal need to revise already-built structure. A left-corner parser is fundamentally a parallel parser, and the ability to parse incrementally (provided by the left-attachment enhancements) does not change this. The need for parallel representations can be seen quite clearly in the example presented above in (36): notice that the relative clause material (the CP and the operator) attached to the noun *fox* is not needed if there is no relative clause. In a serial parser, if the structure in (36) for *the farmer chased the fox* were built for every sentence that starts with that string, the CP would need to be removed in (37) and an extra projection of the verb would need to be added as an attachment site for the locative PP.

(37) The farmer chased the fox out of the hen-house.

Under the assumption that every change in syntactic structure has some cost, effects of this sort of extensive restructuring should be seen, but these sentences remain quite easy to understand.

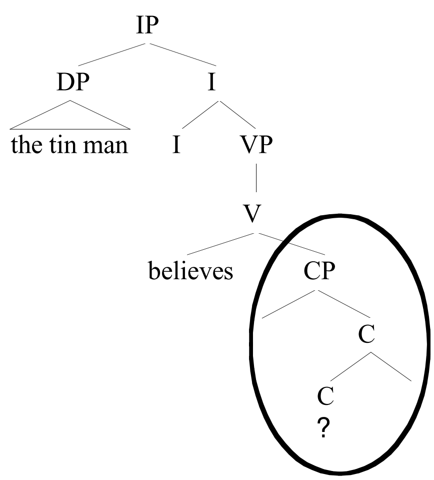
It is crucial to the left-attachment parser that this extra material be predicted before the determiner *the* is encountered if the prediction for a relative clause (and the associated CP) is not made, the determiner that starts the relative clause cannot be attached incrementally. Thus, in order to ever be able to attach a relative clause, a left-attachment parser must predict a relative clause for every NP.

Another case in which a substantial amount of (possibly unnecessary) extra structure must be predicted by the left attachment parser is that of sentential complements.

(38) The tin man believes the scarecrow wants a brain.

If this sentence is to be successfully parsed by the LA parser with a sentential complement to *believes*, the circled structure shown in (39) must already be built at the point at which the determiner *the* is encountered.

(39)



The circled structure is required so that *the* can be attached immediately. However, there are other continuations that do not involve the null C complement needed for (38), as can be seen in (40)-(42).

(40) The tin man believes the scarecrow to be a fool.

(41) The tin man believes that the scarecrow wants a brain.

(42) The tin man believes the scarecrow.

In any of these sentences, a left-attachment parser would need to retract the predictions that are required to parse (38) incrementally. This need to either maintain many parses in parallel or frequently retract predictions is not in keeping with the desiderata laid out at the beginning of this chapter. Thus, the basic problem is that in order to allow for the significant variety of structures allowed in many temporarily ambiguous situations, a left-attachment parser must either perform an fairly exhaustive parallel parse, or it must contend with very frequent backtracking/repair.

1.2.3 Limits on Structural Change

This section provides a short survey of how parsing breakdown has been accounted for in a number of different theories.

1.2.3.1 Frazier s Garden Path Theory

In Frazier and Fodor s (1978) theory of parsing, any error in parsing was argued to lead to a garden path. Since that time, it has been noticed that there are numerous differences in the costs of errors, depending on which particular error is made. The theories that are discussed below are attempts to show how sentences that cause complete parsing breakdown and those that merely cause parsing difficulty can be differentiated.

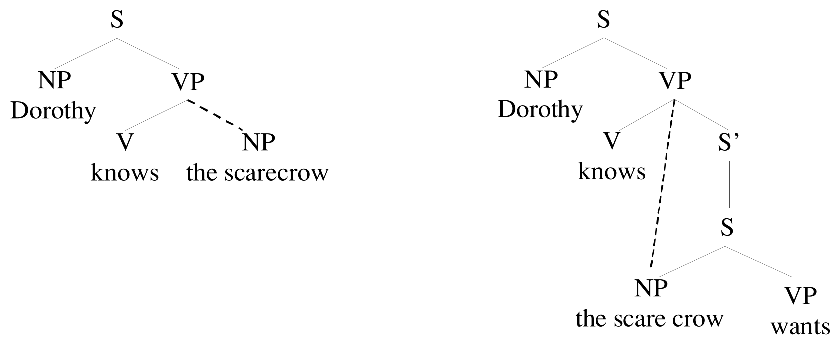
1.2.3.2 D-Theory

Description-theory (Marcus, et al. 1983) is an extension of the Determinism Hypothesis, introduced by Marcus (1980), that is designed to provide an account of parsing breakdown. Description-theory (hereafter called d-theory) is a theory of parsing in which the requirement that trees be built using immediate domination relations (the relation used in traditional phrase structure rules) is weakened. In particular, d-theory allows syntactic representations to be built using domination rather than strict domination. Because syntactic structure is not necessarily represented with strict domination relations, a syntactic node can be lowered in a tree without the need to retract any domination statements. For example, in the sentence in (43), the NP *the scarecrow* can be moved from the direct object position to subject of an embedded clause without the need to retract the domination relation that has been asserted between the VP node and the NP *the scarecrow*.

(43) Dorothy knows the scarecrow wants a brain.

This is illustrated in (44), where solid lines indicate strict domination, and dashed lines indicate simple (non-strict) domination.

(44)



The main premise of d-theory is that structural relations are built monotonically; restructuring that requires the retraction of structural relations (domination statements) is beyond the limits of the parser and should lead to parsing breakdown.

The type of restructuring that is not possible in a d-theoretic parser is exemplified in the well-known garden path sentence in (45), repeated from (6).

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

In this sentence, the NP *the sock* is initially analyzed as the direct object of *mending*. In the grammatical analysis of the sentence, however, *the sock* is the matrix subject. According to d-theory, the problem with this sentence is that the reanalysis from direct object of *mending* to matrix subject is not possible because it involves retracting the domination statement between the VP node associated with *mending*.

D-theory has been the inspiration for a number of theories of monotonic structure-building, including those of Weinberg (1993), Gorrell (1995), and Sturt and Crocker (1996, 1997).

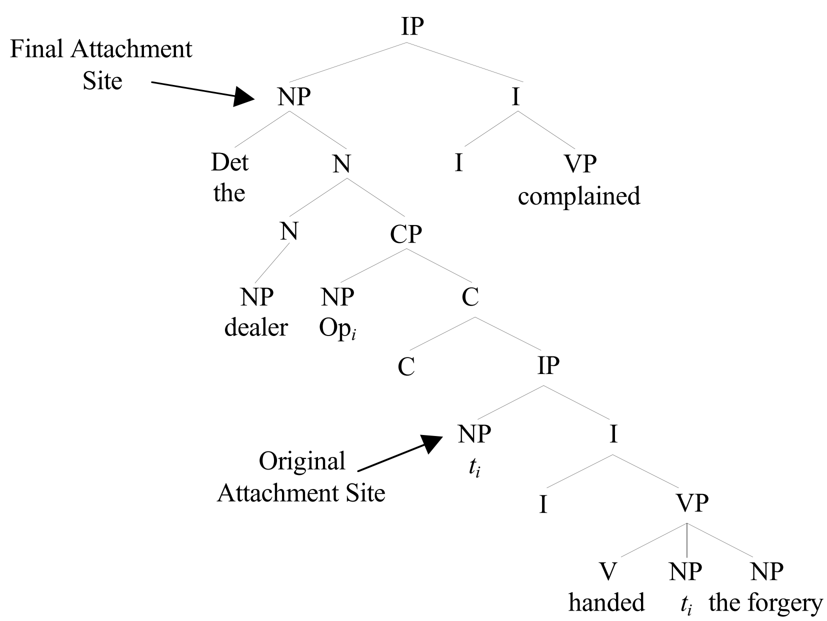
1.2.3.3 Locality Constraints

Pritchett (1992) proposes that there are specific limits on what types of reanalysis are possible. In particular, Pritchett proposes the On-Line Locality Constraint (OLLC), which states that the target position (if any) assumed by a constituent must be *governed* or *dominated* by its source position (if any), otherwise attachment is impossible for the automatic Human Sentence Processor.   Thus, in Pritchett s system, the relation between the original attachment site (the source position) and the final attachment site (the target position) must be either government or domination. Consider the sentence in (46).

(46) The dealer handed the forgery complained.

The difficulty experienced at the verb *complained* is accounted for by the OLLC in the following manner: the original attachment of *the dealer* is as the subject of *handed*. However, as can be seen in (47), the final attachment site (subject of *complained*) is neither governed nor dominated by the subject position of *handed*. Thus, this sentence is correctly predicted to cause parsing breakdown.

(47)



1.2.3.4 Diagnosis and Cure

Fodor and Inoue (1994, 1998) present a theory in which processing breakdown is predicted in a rather different fashion. They state that no repair processes are intrinsically costly. Repair costs depend entirely on the difficulty of determining what revisions to make. They claim that when no grammatical attachment for a word can be found, an ungrammatical attachment is made (by virtue of their Attach Anyway principle), and the parser then follows a chain of grammatical dependencies in an attempt to repair the structure. In the example in (48), repeated from (45), Fodor and Inoue claim that the difficulty results from the fact that Attach Anyway attaches *fell* as the matrix verb (in spite of the lack of a subject for the verb), and there is no grammatical dependency that relates the matrix verb *fell* to the direct object of the subordinate clause *while Mary was mending*.

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

They claim that reanalysis of *the scarecrow* in (49), repeated from (43) is easy because *Attach Anyway* attaches *wants* as a sister to *the scarecrow*, and the sisterhood relationship therefore allows access between *wants* and *the scarecrow*.

(49) Dorothy knows the scarecrow wants a brain.

Because *the scarecrow* can be accessed from *wants*, it is trivial to move *the scarecrow* into a position in which it is the subject of *wants*.

1.3 Summary

The preceding sections surveyed a variety of psycholinguistically-motivated models of parsing with an eye towards how they achieve incrementality and how they account for parsing breakdown. While a number of models were shown to be capable of significant incrementality, none of them specify how full incrementality can be achieved in both head-initial and head-final languages. In what follows, I will show how slight modifications of Stevenson s idea that heads cannot participate in the parse until they have been unambiguously signaled will allow structure to be built in a method very similar to Stabler s Left Attachment model. The SPARSE model allows for incremental structuring of both head-initial and head-final languages, using structures that do not require extensive retraction of assumptions.

I will also show where this model predicts parsing breakdown. A particular prediction of this model is that breakdown should occur when the constituent that is needed to make an attachment is not accessible in the tree (either because the constituent does not exist or because it is in an inaccessible portion of the tree). This prediction is of a different sort than the predictions of breakdown made by d-theory or the OLLC. Those theories predict breakdown on the basis of an inadmissible configuration between the initial attachment site of a constituent and its final attachment site. In contrast, this prediction is based on the availability/visibility of constituents in their original attachment sites; the eventual attachment site is only relevant in that it is what drives the failed search for a particular constituent. In this sense, the prediction is similar in spirit to the model proposed by Fodor and Inoue, for whom visibility is also the important factor, although visibility is computed differently in their model.

The main objective of this dissertation is to show how human language can be parsed incrementally. In Chapters 2 and 3, I provide a detailed specification of an algorithm for the parsing process, as well as examples of how it accounts for incrementality in head-initial and head-final languages. In Chapter 4 I show how wh-movement can be computed and also show how this system can be constrained to prevent extraction out of certain domains. The intent of Chapter 4 is not to show that constraints on movement have a purely processing explanation (i.e. that grammatical constraints on movement are superfluous), but rather, the goal is to show how grammatical constraints can be translated into a parsing system that will accept as grammatical all and only the movements accepted by the grammar as grammatical. Chapter 4 also provides an example of how the parser can structure ungrammatical input, thereby allowing it to be interpreted, but still retain the knowledge that the input sentence is ungrammatical. Chapter 5 contains the details of two psycholinguistic experiments performed to determine how an incremental parser should search the grammar space for an acceptable attachment for incoming material. Finally, the details and status of the implemented version of the parser are discussed in an appendix.

Chapter 2

STRUCTURE BUILDING

2.1 Introduction and Background

This chapter provides a detailed account of how syntactic structure is built in this parser. As has been mentioned previously, the goal of this parser is to build up connected syntactic structures word-by-word in a serial fashion, while minimizing the need to revise structure. The main insight to be gained from this chapter (and the following chapter on head-final languages) is that by building only the structure than can be reliably predicted, there is no need to later reject the structural assumptions, as the structure is compatible with multiple interpretations. At the same time, this structure is able to serve as the basis of incremental interpretive processes that crucially require syntactic structure. Thus, this method will be shown to balance the need for structural commitment and the need to retain flexibility. Structure is projected up on the basis of the incoming word, and because it is projected up only on the basis of confirmed material, it can be certain that the projected structure will not need to be retracted later in the parse. Structure is projected on the basis of relations that the incoming word will necessarily be involved in (e.g. every NP must be assigned case), thus assuring that any predictions made will be confirmed in the final structure (e.g. that the head that takes an accusative NP as its complement will also assign accusative case).

In many other parsing systems, structure is predicted based on the selectional possibilities of a word. For example, in order to incrementally structure an ambiguous verb like *know* (which takes either an S or an NP complement) with a following determiner and allow the determiner to be part of an embedded clause, it is necessary to predict that CP and IP nodes are present between the verb and the determiner. In a parser that builds structure based on phrase structure rules, this would require use of the following two rules (assuming simple binary branching structure): CP_C IP, IP_NP I . Thus, this sort of system requires that a null C and an I be predicted. However, if the determiner is part of a simple direct object, the predictions of a C and an I will need to be abandoned or revised. A licensing parser would require similar predictions to account for the possibility of a sentential complement, and again, the structures would need to be abandoned or revised if the determiner is actually part of a direct object complement of the verb.[15] The problem with these systems is that if they only predict structure that is guaranteed to follow, they won t predict very much at all, since many categories/heads have significant ambiguity in their selectional requirements.

The question of what structure can be safely predicted is particularly interesting in head-final languages, where much more needs to be predicted than in head-initial languages. Structural predictions must be made so that incremental interpretation can take place. Because the NPs in a string of verbal arguments are not syntactic dependents of one another, some aspect of the verbal structure must be built in advance of the verb, if the NPs are to be interpretable. In head-final languages it may be risky to predict that a particular category will appear, since e.g. an NP could be the sister of either a post-position or a verb. SPARSE gets around this problem by changing the grain of prediction to the level of the feature, instead of the more traditional level of syntactic head (including category). For example, if there is a cased NP, there must be a case-assigner, but the head containing that case-assigning feature need not be immediately specified for a particular syntactic category, since it could end up being either a verb or a postposition.

[15] In fact, in the parser proposed here, a commitment to either a direct object or embedded subject analysis must also be made, and will sometimes need to be revised. This example is meant to be a simple example of predictions that must be made. See Chapter 3 for examples in which this parser preserves ambiguity that other parsers are not able to preserve.

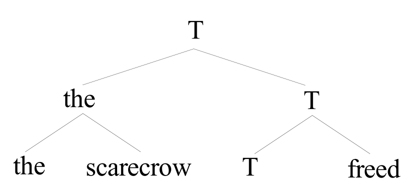
2.2 The SPARSE Model: An Overview

In SPARSE, structure is built bottom-up as each word is received in the input. Only one structure is built and maintained from word to word, and parsing is driven by the need to maintain a single parse tree. Once a single tree has been obtained, there is no need to do further syntactic processing until the next word appears. The only information that is maintained from word to word is the information contained in the syntactic tree, which is maintained so that semantic interpretation can take place. The fact that SPARSE retains only the information in a single tree differs from many other parsers that keep information about possible revisions/choice points (as in backtracking parsers and the variant thereof in Inoue and Fodor 1995) or about dispreferred parses (as in parallel parsers). When incoming words cannot be straightforwardly integrated into the existing tree, backtracking to undo previous decisions is not an option, since that would require a memory store to show how the parse tree was built up. When there is a problem with accommodating new input, the structure is revised through a process of repair that involves many of the processes used in normal structure building rather than by re-parsing the previous input.

In order to retain the ability to easily accommodate multiple possible continuations of an input string, no structure is pre-built. Rather, structure is built as evidence appears in the input. For example, when *freed* is encountered in (50), nothing other than the material needed to connect the NP *the scarecrow* and the incoming verb *freed* is built. The tree in (51) shows the structure after *freed* has been parsed.

(50) The scarecrow freed the lion.

(51)



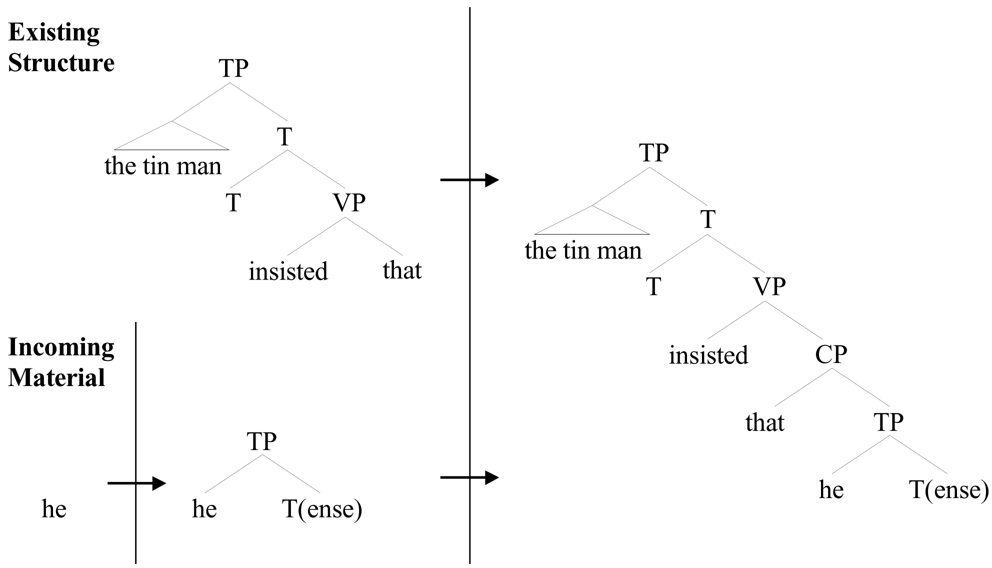
While a number of (serial) parsers would posit an NP node representing the complement of the obligatorily transitive verb *freed* (e.g. Gorrell 1995, Crocker 1996, but not Pritchett 1992), that structure is not built in this parser. The NP is not built, because it is not needed to combine *the scarecrow* and *freed* into one tree. While the object NP can be predicted with certainty, the extra nodes are not needed for the interpretation of *the scarecrow freed*, and thus they are not built. In systems where the NP is posited immediately, it is normally done to show that an NP complement is required of the verb. In SPARSE, this information is carried independently on the verb head itself (as it presumably is in any other system), so an NP does not need to be built to show the combinatory possibilities of *freed*. In SPARSE, a DP node following *freed* in (50) is not posited until the determiner *the* is received. If the sentence ends before an obligatory complement has been received, the licensing features that require a complement will not have been satisfied, so the sentence will be ungrammatical.

The processing of the IP in the subordinate clause of (52) provides a good illustration of what happens when there is not enough structure to connect the new input and the existing tree.

(52) The tin man insisted that he needed a heart.

The problem is that *he* cannot be attached directly to the complementizer *that,* because the features of *that* require a tense head, and therefore do not allow the attachment of the pronoun *he*. To solve this problem, extra material is added to the incoming word *he* on the basis of the features of *he*, as seen in (53).

(53)



Exactly how features are used to build new structure will be discussed in ⁄ 2.5.2 below. In this case, the only head that *he* could be attached to is a tense head, because tense heads are the only heads that assign the necessary nominative case. Accordingly, a tense head is built and attached to the NP *he*. Once a TP is present, the new structure can be added to the existing tree, since a TP is a permissible complement for a C. Because the actual tense head has not yet been seen, it is marked with a   to indicate that it is a predicted head.

2.3 Syntactic Structure

Before discussing in detail how structure is built, I will first specify what type of syntactic structure this parser builds. I assume that the parser builds bare phrase structure, as outlined by Chomsky (1995a,b). In essence, bare phrase structure is a fully lexicalized phrase structure grammar. The most important feature of bare phrase structure for the purposes of this dissertation is that higher projections of heads do not differ in feature content from the heads themselves the projections are completely dependent on the head for all syntactic features. I will briefly sketch below why this is important more details of its importance will be discussed in the course of this chapter.

Another important aspect of bare phrase structure is the absence of vacuous projections. The advantages of this will be discussed later in the chapter.

Instead of having its own features, a higher projection of a head (the equivalent of X and XP nodes) contains a pointer back to the head it is projected from. Whenever the features of a higher projection (e.g. a VP) need to be accessed, they are read directly from the head itself, as suggested in Chomsky (1995a,b). One advantage of projecting everything directly up from the head (and not even having a separate representation of features in the higher nodes) is that the information within the head can modified without any need to independently revise the content of the nodes that are projected up from the head. If features are distributed among different nodes of a projection, as in HPSG (Pollard and Sag 1994) or Stevenson s (1994) GB-parser, then when a feature needs to be changed for the entire projection, the features of each of the projections of the head need to be changed in addition to the features of the head itself. In contrast, in a BPS tree, only the feature on the head needs to be modified.

Because higher levels of structure are merely projections of head words, BPS trees are normally drawn with no distinction between heads and their higher projections (i.e. the word itself is used as the label for all projections of the head). I will generally follow this practice, though in some cases I will show the maximal projection as an XP to facilitate reading of the trees. This is done only for the convenience of the reader and does not mean that I am assuming X -style phrase structure for those constituents.

Another difference between BPS and X -theory is that an additional level of structure is projected in BPS if and only if it is needed allow a new constituency relation. Unlike X -theory (Lyons 1968, Chomsky 1970, Jackendoff 1977, Chomsky 1981) in which an X template including both X and X (XP) levels is projected from every head, in BPS there is no X -template to force vacuous (non-branching) projections. Instead, additional levels are only projected in order to accommodate additional structural relations. Thus, heads are the only non-branching nodes. This means that a higher projection of a head is created only when another constituent is merged into that projection. For example, when the a complementizer *that* is used in a sentence, it would immediately project C and CP nodes under X -theory, even if nothing is attached to the complementizer. This contrasts with BPS where the complementizer has no higher projections until they are needed. Assuming that there is nothing in the spec position of *that*, the complementizer will only project up one level, and that won t be done until the IP complement is encountered. While the question of whether or not a full X -template is used will not play a substantial role in this thesis, I note it here because the trees shown here will not contain any extra levels (i.e. no vacuous structure).

2.3.1 Feature as Minimal Unit

It is apparent that the human parser has no particular difficulty parsing head-final languages (see papers in Mazuka and Itoh 1995 for references on Japanese and Konieczny 1996, Konieczny, et al. 1997, and Bader and Lasser 1994 for evidence from German), despite the fact that they contain many temporary ambiguities. Incremental serial models generally have difficulty accounting for the lack of difficulty in temporarily ambiguous head-final structures, since these models generally commit to a single analysis and are forced to backtrack/repair to accommodate other analyses. For example, a sequence of two NPs (nominative followed by accusative) is compatible with both of the continuations in (55) and (56) (respectively, a transitive verb and the combination of a postposition and a verb), and a parser that commits to a single analysis would be forced to backtrack to accommodate the other.[16]

(55) dass er den Weg sieht
that he the path_{ACC} sees
that he sees the path

(56) dass er den Weg entlang geht
that he the path along walks
that he walks along the path

---

[16] See ⁄3.2 for more examples of temporary ambiguity in head-final languages.

To more easily account for this lack of difficulty, I propose to make syntactic structure more flexible by dropping the traditional notion that the minimal building block the parser can manipulate is the head. Instead, the minimal building block accessible to the parser is the syntactic feature, e.g. [Case: DATIVE].

The switch from full head to feature provides a natural way to underspecify the heads that must be posited in order to incrementally parse head-final structures. The use of features as primitives allows incremental structuring of head-final structures but does not require the pervasive retraction of predictions that is characteristic of many other incremental parsers when they are faced with an unexpected completion. In particular, when a head must be predicted before it has been encountered in the input, underspecification allows the parser to posit only the features that can be guaranteed to be part of the eventual head. For example, if a dative NP needs a head to license it, it can be guaranteed that the head that licenses the NP will assign dative case, even though the head could be either a postposition or a verb. To license a dative NP, the parser need only assert that the licensing head contains the feature [Case: DAT, Left], which simply says that the licensing head requires a complement with the feature [Case: DAT] to its left. The categorial identity of the head can be added when the head is encountered or when some other unambiguous clue is reached. See ⁄2.4.1 and Chapter 3 below for more on how licensing heads are used in parsing.

As in most other theories, I consider a head to be nothing more than a bundle of features In this sense, the change that I am proposing is not a radical proposal, but instead a natural extension of existing theories. Where I depart more from standard theories is in what features need to be part of a syntactic node. It is assumed in most other theories that categorial information is an integral part of every syntactic node, and that there is no syntactic node without a category. I will show that a number of benefits follow from dropping this requirement, including the ability to incrementally parse head-final languages without requiring frequent retraction of predictions.

2.4 Structure of Heads

As just mentioned, heads in SPARSE are nothing more than bundles of features. I assume two types of features, those that allow a head to combine with other heads (licensing features , such as [Case, DAT, Left]), and features denoting properties intrinsic to the heads themselves (inherent features such as [Case: DATIVE]).

2.4.1 Licensing Features

In standard context-free grammar parsers, phrase structure rules form the basis for building syntactic structure. In a standard bottom-up CFG parser, when all of the constituents in the expansion of a rule are present in the input, a rule fires and a new constituent is built up out of the existing constituents. The head of the new constituent (i.e. the label of the new constituent) is specified in the phrase structure rule. For example, given the rule VP ″ V NP PP, a VP will be built whenever a V, NP, and PP are seen consecutively in the input.

I do not adopt a system with these types of phrase structure rules, but instead adopt a strategy of using lexically specified licensing features, similar to those of Abney (1989) and Frank (1992). Licensing features represent relations like case assignment, complement selection, and theta-role assignment. Examples of licensing features and how they should be interpreted can be seen in (57) and (58) below.

(57) [Case: ACC, Left]
Accusative case must be found on the head to the left of this feature.
(for a German or Japanese transitive verb)

(58) [Category: NOUN, Right]
A noun must be to the right of the head containing this feature.
(for an English determiner)[17]

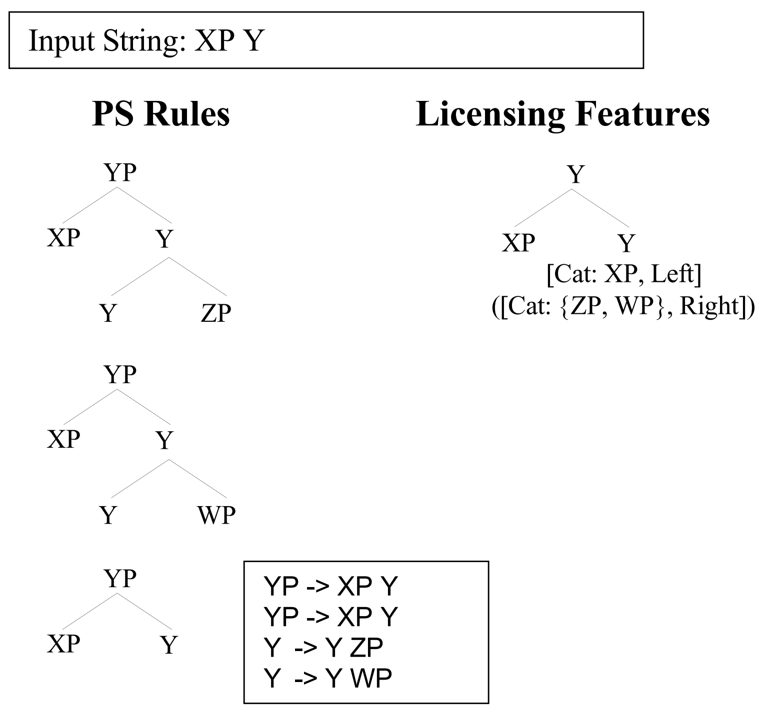
---

[17] Given that the previous section discusses reasons why categories need not be part of every syntactic node, it might be surprising that some heads select for particular categories. However, recall that the claim is that categories need not be part of *every* syntactic node at *every* point in the parse. I do not want to claim that the notion of a syntactic category is entirely irrelevant, but only that syntactic categories are not all-important throughout syntactic derivations.

As can be seen, each licensing feature contains the name of the feature, possible values, and the direction of licensing.

Licensing features are particularly useful in a system that tries to minimize revision/repair as much as possible. In an incremental system that relies on phrase structure rules (e.g. a bottom-up chart parser), a decision about which rule to use must be made as soon the head is used. This effectively means that a decision about what type of complement the head will have must be made immediately. For example, to connect a verb to its specifier, a decision must be made about what sort of complement will be there, if any at all. If there is no complement, a rule that combines a verb and a specifier can be used. If there is a complement, a different rule must be used to allow for an intermediate projection of the verb, which is necessary so that the complement will have a place to attach.

(59)



As can be seen from the example above, if a head Y has three different possible complements (none, a WP, or a ZP), there are three different phrase structure trees that might need to be constructed for the sequence XP Y. Because a phrase structure rule is inherently specified for all of its parts, it is not possible to defer the decision about what type of complement will be needed. In contrast, if licensing features are used to determine attachments, only one structure is needed for the sequence, and the decision about what sort of complement will be used can be deferred until evidence is available to help make the decision. In the licensing features used above, parentheses indicate an optional feature, and curly brackets { } indicate that any of the values of the set may be chosen.

In many cases, a number of relationships must hold simultaneously between a head and a single licensed element, e.g. an English auxiliary not only assigns nominative case to the subject, but it also agrees in person and number with the subject. This is represented by multiple licensing features in a single head, organized in a set. All of the licensing features in a feature set must be licensed by a single head if the licensing relationship is to succeed.[18] For example, a German determiner must agree with its complement (the noun) in number, gender, and case. An example of the set of features needed for the determiner *dem* (dative, 3rd person, singular, masculine or neuter) is shown in (60) below.

(60) {[Person: 3, Right], [Number: S, Right], [Case: DAT, Right],
[Gender: {MASC, NEUT}, Right]}

Licensing features are used in a checking relation similar, but not identical, to that of Chomsky (1995b). In Chomsky s conception of the checking relation, the value of a feature on the checking head (licensing feature on the licensing head) and the checked head (the inherent feature on the head to be licensed) must be exactly the same. This is easily accomplished in his system because elements in syntax are never ambiguous. However, because the features available to the parser can have multiple values (due to

---

[18] As will be discussed below, the features that specify head-complement and head-specifier relationships belong to different feature sets.

lexical ambiguity), the checking relation used here is somewhat different. The basic idea is that the values of the licensing feature and the values in the corresponding feature of the licensed head must have a non-null intersection. If each feature has only one value (i.e. each feature is unambiguously specified for a single value), then the values must be the same, and if there are multiple values, there must be at least one value in common between the licensing feature (on the licensing head) and the inherent feature (on the head to be licensed). If either head has a null value for the feature, the intersection is null and the intersection therefore fails, as does the licensing relation.

Intersection, as used here, is very similar to the concept of unification familiar from computer science, but the two are not exactly the same. The main difference is that unification is successful if the two values to unify are a null value and an actual, instantiated value. Thus, a feature like [Case: ACC, Left] would unify with a head that has no Case feature at all, since the unification of a null value and an instantiated value is the instantiated value. However, this behavior is undesirable in this system, because, for example, a case-assigner should not license a head that has no case feature.

SPARSE contains an explicit variable that can be used as a feature value. When a variable is used in a checking relation, it transfers the values from the other feature in the checking relation. Variables are used for features that can take any value, such as the theta-role feature on an NP. Because there is no restriction on what theta role NPs can take, the variable allows the NP to take any of the possible theta roles. The difference between a variable and a null value in this system is that the variable matches anything, while a null value matches nothing. An alternative to using a variable would be to enumerate all possible values for a feature (e.g. [Theta: Agent, Theme, Goal, Experiencer,]). These two options (variables or full enumeration) are notational variants, and I choose to use variables because it is simpler and more naturally captures the idea that there are no restrictions for some features (e.g. theta-roles on an NP). Features that are ambiguous, but not completely free (e.g. case for the German determiner *den* is restricted to dative or accusative) are represented by a simple enumeration of the possible values.

2.4.2 Inherent Features

The structure of an individual inherent feature is very similar to that of a licensing feature, except that there is no expression of direction of licensing. An inherent feature contains nothing more than a feature label and a set of values, e.g. [Case: ACC]. Like licensing features, an inherent feature may have a single value, multiple values, or it may have a variable as its value. While a single head can have multiple sets of licensing features (for different items it selects, e.g. complement and specifier positions), a head only contains a single group of inherent features.

The difference between inherent features and licensing features is very similar to the difference between the GIVE and NEED relations used in Frank (1992). In Frank s Generalized Licensing theory, a GIVE relation is a relation that determines which other types of lexical items are licensed by the element containing the GIVE. NEEDS specify the licensing requirements of a node. Thus, a node that NEEDS case requires that some other node be able to GIVE that node syntactic case. One difference between Frank s Generalized Licensing and the theory presented here is that Generalized Licensing also requires an additional set of traditional features (similar to the inherent features use in this work) to represent agreement features, categorial information, and other similar information. In Generalized Licensing, adjunction consists of a single GIVE that has no corresponding NEED. For example a VP-adjunct needs to combine with a VP, but VPs do not require adjuncts. Thus, there must be some feature on a VP that the GIVE can target, but which does not function as a NEED. If that feature were a GIVE, the adjunct would be required.

This need for features in addition to GIVES and NEEDS is eliminated in SPARSE by making a distinction between different types of licensing features. To allow for adjuncts, some licensing features are allowed to forgo checking off of the inherent features of the head that they license. Thus, an English VP adjunct in SPARSE has a licensing feature that selects for a head of category V to its left, but does not check off the category feature of the verb. This is necessary because the [Cat: V] feature of the verb will be checked off by the head that it is a complement of (e.g. a Tense head).[19]

2.5 Structure Building

The basic strategy for building new structure, as discussed above, is to attach new material onto the leading edge (the right-most elements) of the current phrase marker. Syntactic attachments are allowed on the basis of licensing features (discussed above in ⁄2.4.1). The attachment operation itself is similar to the Merge operation discussed in Chomsky (1995b), except that here it does not need to apply at the root of the tree.

2.5.1 Simple Attachment

To see how the parser builds structure, it will be useful to work through a simple example. Consider the sentence in (61) below.

(61) Dorothy will see munchkins tomorrow.

When *Dorothy* is encountered, nothing needs to be done, since the head *Dorothy* is a fully connected structure on its own and can therefore be given a semantic interpretation. The next word, *will,* contains two feature licensing sets, with the following among their contents:

(62) will:
{[Case: NOM, Left], [Num: {SG, PL}, Left], [Person: {1,2,3}, Left]}

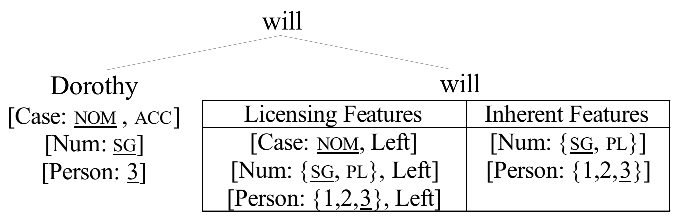
{[Category: VERB, Right], [VForm: INFIN, Right]}

The licensing features are inspected to determine whether either set can license the most recent word, *Dorothy*. One of the feature sets (the first one in (62)) licenses the features of *Dorothy*, so an attachment of *Dorothy* to *will* is carried out. Because the auxiliary *will* licenses *Dorothy*, *will* projects up as the head of the new constituent, as

[19] As will be discussed below, I assume that selectional features (e.g. [Cat: V, Right]) participate in exactly the same sort of checking operations that are used for other features (e.g. agreement features).

illustrated in (63). Note that only the features relevant to each example are included in the trees there are many more features for each head that are omitted for clarity.

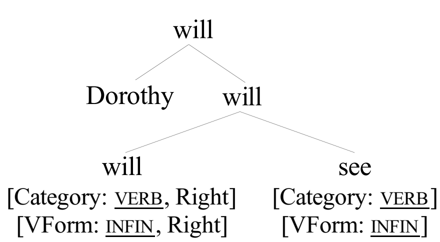
(63)



As you can see in this tree, the auxiliary *will* has two different sets of features. The set on the left is the set of licensing features relevant to this example, and the features on the right are the inherent features of the head. Features that have been checked are marked with underlining under the feature value that was in the intersection. Feature values are also underlined when the value has been further specified. When a head contains an inherent feature and a licensing feature of the same category, the values must always be the same. In this way, the inherent features can be further specified, and can agree with whatever the head is in a checking relation with. For example, the value 3 is underlined in the inherent Person feature on *will* because it has been further specified by the Person licensing feature on the head.

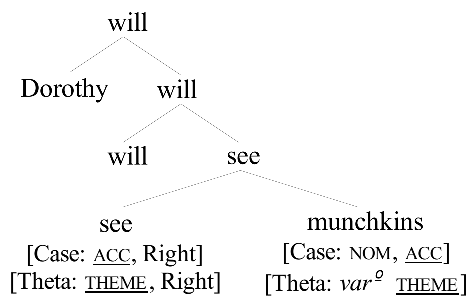
When *see* is encountered, the parser attempts to attach it to the most recent word, in this case the auxiliary *will*. Because *see* contains both a verb category specification and infinitival verb form specification, it is licensed as the complement of *will*. The result of this is seen below:

(64)



Next, consider what happens when the noun *munchkins* is encountered.

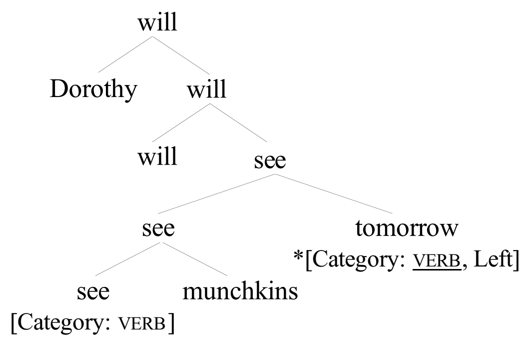
(65)



On the verb, the licensing feature [Case: ACC, Right] can be intersected with the inherent feature [Case: NOM, ACC] found on the incoming noun *munchkins*. Likewise, the value in the licensing theta feature, *theme*, intersects with the variable that is the value of the theta feature for *munchkins*. All of the licensing features can be successfully checked, the attachment is licensed, and *munchkins* is attached as the complement of the verb. At this point, the theta variable in *munchkins* is instantiated with the value taken from the licensing feature of *see* (THEME), and the feature is marked as a checked feature.

In each of the cases we have seen so far, the new word has been attached to the most recently processed word (i.e. the right-branching structure has been expanded). However, in many cases this is not possible. When the attachment site is not the most recent word, the rest of the tree must be searched for an attachment site. When *tomorrow* is encountered, it cannot enter into any licensing relationship with *munchkins*. The next step is to search further up the tree along the leading edge. The next constituent that is encountered is headed by *see*, so the features of *see* are checked against the features on *tomorrow*, with the result that a licensing relation is allowed. Note that in this case the head with the licensing features, *tomorrow*, is an adjunct and therefore does not project up to head the new constituent, in contrast to specifier/complement relations where the head that contains the licensing features also heads the new projection. Adjunct status on licensing features is denoted by an asterisk before the feature, as can be seen in the tree in (66), which depicts the final structure after *tomorrow* has been attached. The fact that this is an adjunct licensing relation must be marked in the lexical entry so that *see* will project up instead of *tomorrow.*

(66)



2.5.2 Predicting Structure

Unlike all of the attachments that were made in the previous sentence, it is frequently the case that when a new word is processed, there are no attachments involving a direct syntactic relationship between an element in the existing tree and the incoming word. A simple example of this was discussed above for (52), repeated here as (67).

(67) The tin man insisted that he needed a heart.

Recall that when *he* is encountered, it cannot be attached directly to the complementizer *that*. Since no direct attachment is possible, a new head is predicted and attached to the incoming word. The question of which features need to be present on the licensing head now becomes important. For example, the features of English wh-words are not all licensed in a single location the wh-features are licensed in [Spec, CP], while case and theta features are satisfied at positions lower in the tree. The feature that must be licensed in the surface position of a pronoun is the case feature; in this example the feature is [Case: NOM]. Because this feature is the one that must be licensed first, the case feature on the pronoun will be distinguished from the rest of the inherent features. This demonstrates a general property of the parser: a predicted head does not need to license all of the features on the word it is attached to. Instead, it merely needs to license the distinguished feature.

When a new head needs to be predicted, the distinguished inherent feature is used for a search of the lexicon. The search returns all the heads in the lexicon that license the feature. In this case, each of the heads that are returned assigns nominative case. The head that is predicted is the head whose features consist exactly of the intersection of the features of each head that licenses the feature [Case: NOM].[20] All of the relevant heads in English also contain a tense feature, so Tense is part of the predicted head. Note that this tense head is compatible with all of the different possible tense heads that could follow *he,* because it is nothing more than the intersection of all of those heads. Once a predicted tense head has been attached to *he*, the *he*-T complex can be attached to the complementizer *that*, by virtue of the tense-selecting feature on the complementizer.

Next, consider what happens in the more complicated case of the following sentence, where the complementizer is missing and the case value of the subject is ambiguous:

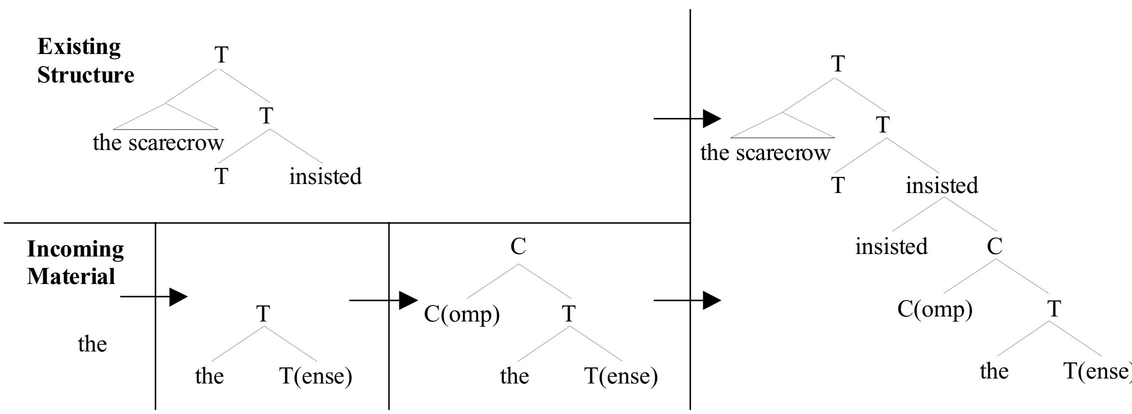
(68) The scarecrow insisted the tin man needed a heart.

When the determiner *the* is encountered, it cannot be attached directly to the existing tree, because *insisted* does not assign case. Because direct attachment is not possible, a new head is predicted and attached to the determiner on the basis of the inherent features of *the*. I assume a DP structure in which the determiner is the head of nominals, and the noun itself is the complement of the determiner (c.f. Abney 1987). Because the case feature is the distinguished feature for *the,* the predicted head must have a case-assigning feature. However, because the determiner is case-ambiguous, it is unclear which value for case the licenser should have. Fortunately, there is additional information available that can be successfully utilized to solve this problem. Namely, the head that is being posited must either follow the determiner in the input string or it must be phonetically null (since any non-null head to the left of the verb would have already been encountered in the input). Given these constraints, the lexicon is searched for either a null case assigner or

[20] This search could potentially return thousands of heads that contain the relevant feature. Presumably some other process is available to avoid the need to compute the intersection of all of these heads on-line. For example, an associative network might store for quick retrieval the fact that the predicted that licenses a nominative NP also contains a tense feature.

for a head that can assign case to its left. There is no null element in English that can assign case to its right, but Tense heads (the same group mentioned above with (67)) do assign nominative case to their left. Accordingly, a Tense head is built and *the* is attached to it. Even with the addition of the Tense head, there is still no way for [*the* T] to be attached to *the scarecrow insisted*, so another search is initiated for an element that can license the distinguished feature (Tense) of the just-built Tense head. A null complementizer is returned by this search as the unique solution, so a complementizer is attached to the Tense head, forming a new CP constituent. At this point, the entire CP structure can be attached to *insisted* in the existing tree. As previously mentioned, heads that are predicted (i.e. those whose head positions are to the right of the most recently processed word) are marked as such with a   on the head. The entire process of parsing the word *the* in this sentence is illustrated below:

(69)



Note that the CP-selecting property of *insisted* is not used to guide the building of the CP above *the*. There are two reasons for this. The first is that the CP can be predicted independently of *insisted*, since the only head that can license a finite Tense head is a complementizer. Secondly, using information from the existing tree opens up the question of which heads should be involved in structural predictions. In this case it is relatively clear that if something is going to be attached, it should be attached to *insisted*. However, if there are other potential licensers above the most recent word, the parser would have to decide which one to work from. By only building from the new word, this potential problem is avoided.

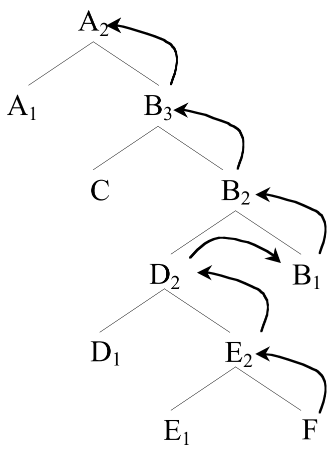
2.6 Logic of Parser

In this section I begin to outline the logic of the parser, filling in many of the details that were omitted above. Recall that main goal of this work is to provide an explicit account of incremental structure-building in both head-initial and head-final languages. In addition to this, I have also attempted to incorporate a number of structural preferences from the psycholinguistic literature. Among the parsing preferences that are incorporated into the algorithm are the following: argument attachment preferences (as discussed thoroughly in Pritchett 1992, Gibson 1991 and Sch tze and Gibson 1999), recency effects which show that it is generally preferred to attach new elements to recently processed structure (see discussion above in ⁄ 1.2.1 and below in ⁄ 5.2.1), the relative ease of reanalysis in many situations (e.g. English NP-S ambiguities, temporarily ambiguous head-final structures), the preference to not reanalyze (summarized in ⁄ 2.7 and fully discussed in Chapter 5), and the extreme difficulty/inability to reanalyze in some situations (e.g. the garden paths in ⁄ 1.1).

In what follows, I will frequently refer to searching the tree. This means searching a tree in a depth-first left-to-right manner starting at the most recently processed word, as illustrated in (70).[21]

[21] Node B_2 is not searched immediately after D_2 because anything that is added at B_2 should be added to the right of the terminal B_1. Thus, B_2 is not searched until after B_1 has been searched.

(70)



By starting the search at the most recently processed word (*F* in (70)) rather than at the root, only the right edge of the tree will be searched, and there will be no attempts to attach an incoming word to the left of any of the already-received material.

The search of the tree stops when any of the following types of heads is reached: the root of the tree, a predicted head, or a head/constituent still missing a required element on its right (e.g. an obligatorily transitive verb that has not yet received its complement, which would be represented by an unchecked right -pointing feature). If any such head is reached before an attachment site is found, the search fails. The search must stop at the root for obvious reasons there is nothing else to search. It must stop upon reaching a predicted head because the predicted head must be instantiated with a real head. If the incoming word is attached to the right of the predicted head, there is no way for another incoming word to instantiate the predicted head without violating the linear order of the sentence. Likewise, if an obligatory complement is passed up, there is no way to attach a complement later in the derivation without violating the linear order of the input sentence.

The search itself is for either a predicted head that the incoming word can instantiate or a licensed attachment between the incoming word and a constituent in the existing tree. The licensing relation can be from the new word to the existing tree or from the existing tree to the new word. If a search is successful, the licensed attachment is made and the parsing for that word is complete. If any incoming word can instantiate a predicted head, the features of the incoming word are integrated into the features of the predicted head and parsing is complete. Whether a word can instantiate a predicted head is determined by whether or not the features of the predicted head can subsume the features of the incoming word. This is discussed in more detail in Chapter 3.

1. Search unchecked features on the right edge of the existing structure for an argument attachment for the incoming material.
2. Search the right edge of the existing structure for an adjunct attachment of the new material.
3. Build a new licenser for the new material:
   - 3.1. Search the lexicon for all possible licensers of the new material. A possible licenser is either a head with appropriate left-pointing features or a null head with appropriate right-pointing features.
   - 3.2. Attach to the new material the intersection of all heads returned by the lexicon search.
   - 3.3. Return to step 1 with the new constituent built in step 3.2.

Figure 1: SPARSE Parsing Algorithm (Preliminary)

This version of the algorithm will be expanded as the following chapters unfold.

2.7 Experimental Results

As mentioned above in the introduction to the parsing algorithm, this parser is designed to account both for recency/locality effects and also for ease of reanalysis in many situations. Reanalysis is exemplified in the following example of the standard direct object NP/embedded subject ambiguity:

(71) The scarecrow knows the tin man wants a heart.

When *the tin man* is parsed, it can be either a direct object of *knows* (as would be the case if the sentence ended immediately after *the tin man*), or it can be part of a sentential subject, as is required for the sentence in (71). In these types of sentences, it has been observed (Frazier and Rayner 1982, Rayner and Frazier 1987, Ferreira and Henderson

1990, but see Mitchell and Holmes 1985 and Trueswell, et al. 1993 for qualifications) that the NP *the tin man* is initially analyzed as a direct object.[22] However, when the verb *wants* is processed, the role of the NP *the tin man* in the sentence changes from that of direct object to that of subject of an embedded clause. This change in roles is accompanied by a change in the syntactic structure. Researchers going back as far as Fodor and Frazier (1980) have suggested that the reanalysis required in these sentences should only be performed as a last resort. They call this effect Reanalysis as a Last Resort (RALR). The idea is that even though the change is a relatively easy one to make, it is avoided unless there are no other options.

The recency/locality preference for incoming material alluded to above has been shown for a variety of constructions (see e.g. Wanner 1980, Gibson et al. 1996, but also Cuetos and Mitchell 1988). This preference can be seen in examples like (72), where there is a preference to interpret *last week* as a modifier of the more recent verb *caught* than as a modifier of *heard*.

(72) I heard that Dorothy was caught in a tornado last week

Tension exists between recency and RALR; that tension forms the basis of the experiments in Chapter 5. The basic tension is this: if RALR is followed, local reanalysis should be avoided in favor of more distant analyses without reanalysis. Recency, however, predicts that the local option (reanalysis) should be preferred over any non-local attachment of the verb. In the standard NP-S ambiguity shown in (71), there is no way to tell if RALR affects recency preferences, since the recent reanalysis is the only one allowed. How this tension is resolved has direct implications for the parsing algorithm. If RALR is in effect, the grammar search should check all other options before trying to carry out reanalysis of existing structure. On the other hand, if recency has a strong enough effect that reanalysis is carried out in place of other possible analyses, then the search algorithm should reflect this fact. Thus, the basic question has

---

[22] The facts are actually somewhat more complicated and depend upon the specific choice of embedding verb, e.g. *knows* in (71). See the discussion in ∕5.2.1 for more details.

to do with how the grammatical space is searched is reanalysis explored early in the parsing process, or is it explored only if all other options fail.

A summary of the results from the two experiments performed to test the tension between recency and RALR is provided here; a full discussion can be found in Chapter 5. To test which of these two constraints has priority, the standard NP-S ambiguity was embedded within a subject relative clause, as shown in (73).

(73) The creative woman who knows the funny man wrote some comedy sketches

In this structure, there are two possible analyses for the verb *wrote*. It can either take *the creative woman* as its subject, or it can take *the funny man* as its subject. If recency/locality is a stronger influence on the parser than RALR (i.e. if reanalysis isn t an absolute last resort), *the funny man* should be taken as the subject of *wrote*, since *the funny man* is more recent than *the creative woman.* On the other hand, if reanalysis truly is a last resort, *the creative woman* should be taken as the subject, since that analysis does not require any changes to existing structure.

The results of the two experiments show that distant attachment sites that do not involve reanalysis are preferred over more local/recent attachment sites. Specifically, if the NP *the funny man* is attached as the direct object of *knows*, the following verb *wrote* takes *the creative woman* as its subject, despite the more local NP *the funny man*. The embedding verbs (e.g. *knows* in (73)) were divided into different classes depending on the relative frequency of NP complements and sentential complements. The results for each of the different verb classes are consistent with RALR. Additionally, there were interesting differences between two different subject groups those who scored highly on a test of verbal ability and those who did not score as highly. A full discussion of the experiments and results can be found in Chapter 5.

To allow for the reanalysis needed in (71) above, steps need to be added to the parsing algorithm, and they need to be ordered so that they will only be performed as a last resort. A revised version of the algorithm can be seen below in Figure 2. The steps added to allow for reanalysis are shown in italics.

1. Search unchecked features on the right edge of the existing structure for an argument attachment for the incoming material.
2. Search the right edge of the existing structure for an adjunct attachment of the new material.
3. Build a new licenser for the new material:
   - 3.1. Search the lexicon for all possible licensers of the new material. A possible licenser is either a head with appropriate left-pointing features or a null head with appropriate right-pointing features.
   - 3.2. Attach to the new material the intersection of all heads returned by the lexicon search. *If the intersection is null, no new licensing heads can be predicted continue on to 4; otherwise* return to step 1 with the just-built constituent.
4. *Search all features (checked and unchecked) on the right edge of the existing structure for an argument attachment for the incoming material.*
   - 4.1. *If attachment is found, remove existing element from the tree, attach it to the new material, and start over with newly expanded constituent at step 1.*

Figure 2: SPARSE Parsing Algorithm (Version 2 of 4)

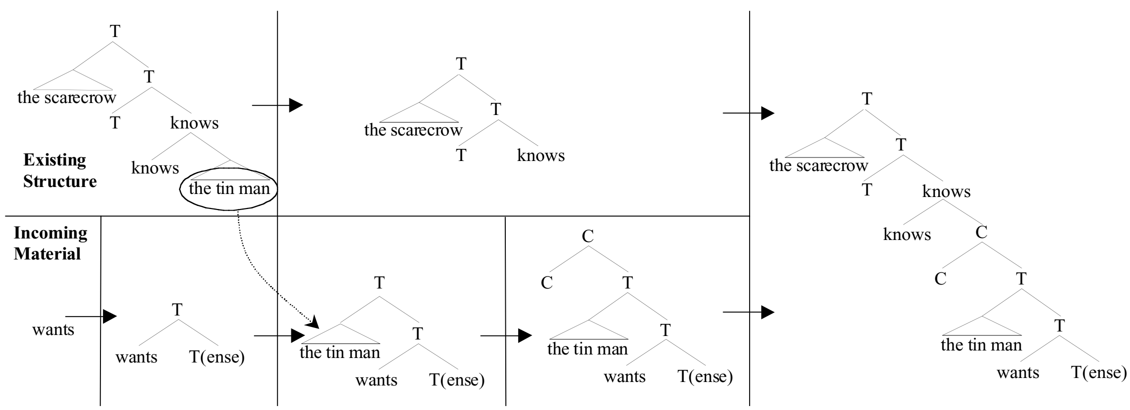
The added rules in steps 4 and 4.1 allow the search to reconsider attachments of nodes that have already been attached. How these additional steps are used is illustrated using sentence (74), repeated from (71).

(74) The scarecrow knows the tin man wants a heart.

When *wants* is encountered, the NP *the tin man* is in the direct object position of *knows*. *Wants* cannot be directly attached into the tree, because there are no heads in the tree that select for a verb. Accordingly, a Tense head is added to the verb, using the same mechanism that has been seen before. This Tense head also cannot be attached into the tree, because nothing in the tree selects for a Tense head (for that matter, nothing in the tree selects for anything at all, since the tree represents a complete sentence without *wants*). A null C head cannot be built to license the T head, because nominative case has not yet been assigned by the tense head. At this point, step 4 is triggered, which allows the parser to consider reattaching an existing node to the new material. Because the T head needs to assign its nominative case, and because the NP *the tin man* can have nominative case, the NP is removed from the existing tree and attached as the subject of *wants*.

At this point, the new word (along with all the other material that has accumulated) starts the basic attachment process again. It still can t be attached into the existing tree, because there are still no heads that can select a TP. The next step is to build more structure above the TP. The only head returned by the lexicon search is a null C, so a null C is attached to the TP. Once the C has been attached, the entire CP can now be attached as complement of *knows.* The entire parsing process for attaching *wants* is shown in (75).

(75)



The basic search strategy and use of predicted heads discussed so far is similar to Konieczny (1996). However, there are a number of differences between SPARSE and Konieczny's system. Among the differences are the grammatical formalism (Konieczny uses HPSG) and the use of predicted heads Konieczny uses fully specified predicted heads. As mentioned above and detailed below in Chapter 3, predicted heads in SPARSE need not be fully specified for all features.

2.8 Predictions

This theory of parsing makes a number of predictions about the processing of different types of sentences. This section outlines the predictions the model makes with regard to both parsing difficulty and parsing preferences.

Within the area of parsing difficulty, it can be seen that in this model there are two different types of difficulty that must be distinguished. The first type of difficulty arises in situations that the parsing algorithm simply isn t able to deal with, e.g. when there is no possible attachment for an incoming word. When the parser is put in such a situation, mechanisms other than the automatic parsing algorithm discussed here must be activated. This will be termed parsing breakdown. [23] It should be possible to observe effects of parsing breakdown in introspection, as well as in on-line measures, such as reading time and comprehension level.

The other possible type of processing difficulty comes from increased processing within the algorithm (which is presumably completely automatic). Parsing preferences arise in this system as a by-product of the search process the attachments that are arrived at first in the process are preferred over those that would be reached later in the search (similar to Frazier and Fodor 1978 and Stabler 1993) This sort of difficulty may be measurable in on-line tasks, but it should not give rise to conscious difficulty or to reduced comprehension levels. I will use the term parsing difficulty to denote this sort of difficulty, which contrasts with complete parsing breakdown. In the literature, the term garden path has been used by various authors to denote either of these different types of processing problem. In order to avoid any confusion, I will refer to failure of the algorithm to deal with the input as parsing breakdown and effects of increased processing within the capacity of the algorithm as parsing difficulty. Note that I do not assume that all operations of the automatic parser are necessarily completely cost-free.

[23] The meaning of this term is similar to the meaning of conscious garden path.

While it is possible that all operations of the automatic parser are cost-free, the variations in the amount of processing required to parse a sentence within the automatic parser are large enough that the differences in amount of processing required by the algorithm for different structures are probably detectable in on-line measures. For example, assuming that there are some costs associated with operations performed by the automatic parser, words that can be attached directly into the existing structure should be processed more quickly than words which require predicted heads and reanalysis.

As mentioned in ⁄2.6 , a number of parsing preferences discussed in the psycholinguistic literature on ambiguity resolution were incorporated into the design of the parsing algorithm. Among the preferences built into the model is a preference for local attachments over non-local attachments, as long as the two attachments are otherwise equivalent. For example, a preference to attach to more recent material predicts that temporal adjuncts that can be attached to either a closer or more distant VP will generally be attached to the closer VP, as in (76) where *last week* is preferably interpreted as when the tornado struck, rather than when the tornade was discussed.

(76) Auntie Em said that a tornado struck last week.

This preference is instantiated in the model in the search mechanism, which always starts at the most recent word and moves up in the tree from there. Thus, the more local attachment should always be found first.

Argument attachments are predicted to be preferred to adjunct attachments, regardless of locality, assuming that the argument attachment does not require that extra nodes be posited. An example of this can be seen in the sentence in (77), where the difficulty at *in my mouth* is presumably caused by *on the table* being parsed as an argument of the verb *put* and not as a (more local) noun modifier (Gibson 1991, Pritchett 1992, Gorrell 1995, Sturt 1997)

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

In SPARSE, the preference for an argument attachment of *on the table* stems from the fact that the parser searches for all argument attachments before any adjunct attachments are considered. Thus argument attachments are preferred over adjunct attachments, even if the argument attachment is less local than the adjunct attachment. See Sch tze (1995) for a review of the empirical evidence for this claim.

Another preference built into the model is a preference for analyses that require less added structure (i.e. Minimal Attachment). In SPARSE, this preference for less new structure is accounted for by the way in which new structure is built. Because extra structure is built one head at a time, an attachment which requires less extra structure should be encountered earlier in the parsing process than an attachment that requires more predicted structure. Recall that extra structure is built up one head at a time, so a structure requiring no extra heads will be encountered earlier in the search, and therefore be preferred to a structure that requires one predicted head. Likewise, an attachment that requires building one extra head will be preferred over an analysis requiring two extra heads.

Chapter 5 presents evidence showing that lexical biases can cause some ambiguous verbs behave as if they are unambiguous. I assume that this happens because features are sometimes not available to the parser immediately. Because all features are not always available to the parser, Minimal Attachment does not have the same effect in this parser that is has in other parsers. The standard version of Minimal Attachment states that the attachment that requires the fewest new nodes will be preferred. Because of the lexically-modulated availability of features in SPARSE, the reference set for minimal attachment may be changed. In particular, the features necessary for the minimal attachment may not always be available. In these instances, the attachment that is made by SPARSE will be the minimal attachment that is in the set of available attachments. Thus, Minimal Attachment is still in effect (by virtue of the way in which structure is built), but its effects may be obscured by lexical biases.

In addition to the preference for fewer extra heads, this model also predicts that the addition of extra heads may be reflected in parsing difficulty, due to the extra processing necessary for each extra head. For example, the attachment of *she* should be easier in (78) than in (79) because the complementizer head does not need to be posited in (78).

(78) Dorothy insisted that she be allowed to go back to Kansas.

(79) Dorothy insisted she be allowed to go back to Kansas.

Whether this sort of difference is large enough to be measured using current techniques is unclear, but the prediction is quite clear attachments requiring fewer extra heads should be parsed more quickly than those that require more extra heads.

The discussion will now turn from parsing preferences to parsing difficulty/ breakdown. One situation where parsing difficulty, but not breakdown, should be seen is reanalysis in which the incoming word causes a node in the existing tree to be reanalyzed to a different location in the tree. An example of this is the standard NP-S temporary ambiguity in (80), repeated from (71) above. An unambiguous control is seen in (81).

(80) The scarecrow knows the tin man wants a heart.

(81) The scarecrow knows that the tin man wants a heart.

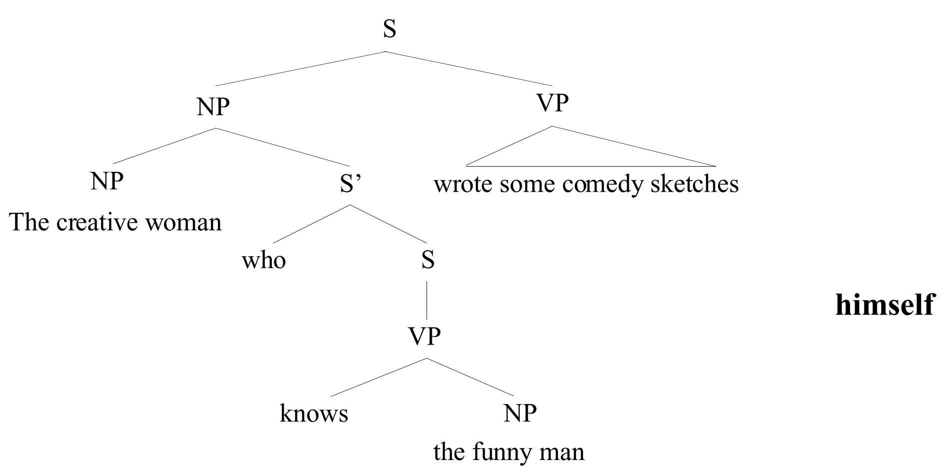
When *wants* is reached, if the NP *the tin man* has been attached as a complement of the preceding verb, it must be moved from direct object of *knows* to subject of *wants*, which is part of an embedded clause. Although the parsing algorithm is capable of doing this, the SPARSE systems predicts that *wants* will be harder to analyze in (80) than (81) because of the amount of work that must be done both to determine what to reanalyze and to carry out the actual reanalysis. In particular, searches (that will eventually fail) must be done for argument and adjunct attachments of *wants*, new structure must be built and checked for possible attachment sites, *the tin man* must be removed from the existing tree and added to the new structure, and finally, the new embedded clause must be attached into the existing tree. This difficulty may be reflected in increased reading times or increased anomaly detection rates (as measured by the stops making sense task of Boland, et al. 1989) at *wants* in (80) in comparison to (81).

The fact that any reanalysis is needed at all in these sentences is related to the prediction that, in general, attachments that can be made by positing less structure (e.g the direct object attachment of *the tin man* to *knows*) are preferred to attachments that require more extra structure (e.g. an embedded subject analysis of *the tin man* after *knows*) to be posited.

Parsing breakdown is predicted for sentences in which new material attaches to existing structure at one location, and thereby forces a change in the structure at a different location. Consider the sentence in (82) (the same as (212) below in the experimental section).

(82) The creative woman who knows the funny man wrote some comedy sketches himself about the amusing escapades thinks he should publish them.

(83)



If reflexives are initially attached into trees without regard to their binding requirements, the reflexive *himself* should attach as part of the VP headed by *wrote*. The experimental results in Chapter 5 suggest that *wrote some comedy sketches* is initially analyzed as the matrix VP (as in (83) below). Thus, if the reflexive is attached to the VP headed by *wrote*, there is no way for the reflexive to be locally bound by a masculine antecedent. Given this situation, the solution should be to make the matrix VP part of an embedded clause inside the relative clause headed by *the creative woman*. However, because the automatic parser can only initiate restructuring as part of attaching a new word, and because *himself* has already been attached, the automatic parser is unable to initiate the change (remember that we are assuming in this example that binding does not play a part in initial attachment decisions). Once a word has been attached, the automatic parser is finished, so some other mechanism must be relied upon to force the change.

In this situation, the parser is unable to initiate the syntactic restructuring necessary because the attachment site is not directly related to the site where the change must occur. This contrasts with the standard NP-S reanalysis discussed with sentence (80). In that sentence, the site of reanalysis is directly affected by the process of attachment needed for the incoming word, because the incoming word *wants* triggers the stealing of the direct object *the tin man*. The differences between these two reanalyses point to a more general prediction of the model: that structural change is generally possible at the site of attachment of a new word, but not at sites unrelated to the incoming word.[24]

[24] The definition of which sites can be affected by reanalysis differs from the reanalysis model presented in Fodor and Inoue (1994, 1998). As noted earlier, restructuring can take place in Fodor and Inoue s model at any site that is related by a grammatical dependency to a syntactic node that is incompatible with its present attachment site. In their model, an attachment of a node X at one site can cause a node Y at a different site to become incompatible with some aspect of its attachment. In turn, this can cause the node Y to force a change at a different node Z. The only requirement is that X be related to Y by some grammatical dependency and that Y be related to Z by some grammatical dependency. There are no distinct constraints on the relation of X (the new word) to Z or any other words that trigger some sort of change in the tree. The only requirement is that there be a chain of grammatical relations between the new word that triggers the changes and the nodes involved in the changes.   In contrast, the SPARSE theory requires that each of the nodes that trigger an attachment change to be at the root of a subtree containing the incoming word.
Another area in which this model differs from Fodor and Inoue s model is that the fundamental operation of the SPARSE parser is attachment. In SPARSE, the only way that a node can be detached from a structure is if the detachment is forced by the need to attach the node to another node (e.g. reanalysis as discussed for NP-S ambiguities). By contrast, Fodor and Inoue allow nodes to simply be expelled from their present attachment site because they are no longer consistent with their present attachment site. They are then subject to normal attachment processes. Thus, the system of Fodor and Inoue allows for more reanalysis than is possible in the SPARSE model.

Parsing breakdown is also predicted in situations in which the grammatical requirement of the new word cannot be met by anything in the existing tree. Consider what happens at *himself* in (82) above if reflexives require antecedents as a condition of attachment. As can be seen from the tree in (83), there is no position on the right edge of the tree (above the most recent word, *sketches*) where *himself* can have a masculine antecedent. The only possible attachment sites are in the matrix VP or the matrix S, neither of which has any possible masculine antecedent. Because the requirements necessary for successful attachment cannot be met, parsing should break down completely. Breakdown occurs in this situation because of the general fact that the parser cannot restructure existing material in order to make a desired piece of structure available. If the desired structure has already been built, it can be used, and even moved if necessary, but there is no way for an incoming word to force the restructuring of the existing tree in order to make available a piece of structure needed for the incoming word. In other words, if there is no masculine subject, the tree cannot be modified so as to make one available.

I know of no existing experimental attempts to determine whether or not the antecedent of a reflexive needs to be computed before it is attached into the syntactic tree. However, in the spirit of this dissertation, I assume that the antecedent of a reflexive needs to be computed before the reflexive can be definitively attached into the existing structure. Thus, I believe that the difficulty with the reanalysis is related to the fact that the reflexive itself cannot be attached, rather than to the fact that the reflexive is attached, but cannot be appropriately bound. I attribute the fact that sentences like (82) do not seem to be as difficult as the sentence in (84), repeated from (5), to the fact that it is easier to determine what needs to be done to allow the troublesome word to be attached into the existing tree in (82).

(84) The horse raced past the barn fell.

When the automatic parser is incapable of building structure, as it is in both (82) and (84), I assume that a process similar to that of Fodor and Inoue (1994, 1998) is responsible for structure-building and interpretation. In the Fodor and Inoue system, the difficulty in understanding these types of structures is directly related to the amount of information provided by the error signal. In their terms, if the problem can be diagnosed easily, it can also be repaired relatively easily. In sentences like (82), the reflexive provides enough information to determine that *the funny man* should be part of the same clause as *himself*. Once this is clear, it is possible (though not necessarily easy) to infer that the matrix VP needs to be lowered into an embedded clause with *the funny man* as its subject. The verb *fell* in (84) does not provide such useful information. Instead, it merely carries the information that it must have a subject, but provides no hint about whether its subject should be *the horse* or *the barn*, and it certainly doesn t give any clues to the crucial fact that *raced* needs to be interpreted as a passive verb in a reduced relative clause.

2.8.1 Non-structural factors

In the predictions presented above I focused on how different structural configurations can lead to differential results in parsing. In other words, one structure might be easier to parse than another because of the way its structure interacts with the parser. However, I do not claim that structural factors are the only factors that enter into whether a sentence is easy or difficult to parse. For example, in Chapter 5 I show that lexical factors (as measured through frequency biases) have a significant effect on the course of parsing (this has also been shown in many other experiments see Chapter 5 for references to that body of work). If a verb that can take either NP or sentential complements is normally used with a sentential complement, subjects seem to prefer to interpret an NP following the verb as the subject of an embedded clausal object of the verb. In contrast, if the verb is normally used with an NP complement, there is a preference to interpret a following NP as a direct object of the verb. These sorts of lexical preferences could be modeled within SPARSE by assuming that the subcategorization feature values can have different strengths. If a given verb occurs with a sentential complement 75% of the time, the parser might interpret this verb as only allowing a sentential complement some proportion of the time. This would require the introduction of a probabilistic mechanism to the model, but I see no serious impediments to such an addition.

A separate body of work has shown that pragmatics and context also have an effect on parsing (see for example Altmann and Steedman 1988, Crain and Steedman 1985, Altmann 1988, and Tanenhaus and Spivey-Knowlton 1996). Crain and Steedman show, for example, that the number of entities in the context can have an effect on how ambiguous structures are interpreted. Consider the sentences in (85) and (86).

(85) The psychologist told the woman that he was having trouble with her husband.

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

Both of these sentences are ambiguous up the point of parsing *with*. In a context in which the psychologist is counseling two different women, the disambiguation in (86), which involves analyzing *that he was having trouble with* as a restrictive relative clause modifying *the woman,* does not cause processing difficulty while the disambiguation in (85), in which *that he was having trouble with* must be analyzed as a complement of *told*, does cause a garden path effect. When the context is changed and the psychologist is only treating one woman, the effect reverses, and (85) can be processed easily, while (86) induces a garden path.

Crain and Steedman claim that their findings are due to the fact that the parser prefers interpretations that violate fewer presuppositions or entailments in the present discourse context. In the context of a single woman, the use of a restrictive relative clause to pick out a single woman introduces the presupposition that there are multiple women in the context, which is clearly at odds with the context. By contrast, in the context of two woman, using the term the woman without any restriction contains the presupposition that only one woman is involved.

However, the theory of Crain and Steedman provides no account of how structure can be identified and constructed. Additionally, it does not account for all sentence processing effects. Consider the garden path sentence in (87).

(87) That coffee tastes terrible surprised John.

As noted in Kurtzman (1985), there is no reason to assume that the reading in which *that* is a demonstrative specifying which coffee carries fewer presuppositions with it than the reading in which it is a complementizer starting a sentential subject. Thus, Crain and Steedman s theory provides no explanation of why the demonstrative reading of *that* should be preferred so strongly over the complementizer reading in this sentence. While this example does not show that the ideas presented by Crain and Steedman are wrong, it does show that they must be augmented.

Another example for which Crain and Steedman s theory provides no explanation is the familiar recency effects on adjuncts. This effect can be seen clearly in (88).

(88) Dorothy said the lion cried loudly.

The manner adverbial *loudly* is normally understood as a modifier of the lower verb *cried*, rather than the less recent *said*. Both of these attachments requires an assertion that something was done loudly (either the saying or the crying), but neither of them requires more extra assertions than the other. Thus, Crain and Steedman s theory provides no explanation for the observed recency effect.

In other cases, it appears that Crain and Steedman s theory makes the wrong prediction. For example, in (89) and (90) (from Gibson 1991), Crain and Steedman predict that the sentence involving a restrictive relative clause ((90)) should cause a garden path because it contains more presuppositions than (89).

(89) Is the block sitting in the box?

(90) Is the block sitting in the box red?

The presupposition associated with the restrictive relative clause in (90), that there are at least two blocks, is in greater conflict with the null context than the sentence in (89), which only presupposed the existence of a single block. If this is true, (90) should cause a garden path in the null context, but it does not. Thus, it appears that other factors are not only present in processing, but are capable of overriding the contextual effects discussed by Crain and Steedman.

Because this theory does not provide an account of incremental structuring, and because it does not account for the parsing effects mentioned above, it seems clear that more is needed. In particular, a theory of initial structure-building, such as the one provided in this dissertation, is clearly needed.

Chapter 3

HEAD-FINAL LANGUAGES

3.1 Introduction

Head-final languages pose a significant problem for strictly bottom-up parsers that attempt to build structure incrementally. In head-initial languages, heads are available to immediately guide the parsing of all complements, but are not available to guide the parsing of specifiers. In head-final languages, the information provided by the licensing head is not available to guide the parsing of either specifiers or complements. The fact that this information is not available to guide the attachment of either specifiers or complements forces certain changes to the parsing algorithm. To get an idea of the problem posed by head-final languages, consider the sequence of Japanese NPs in (91) below:

(91) Yoko-ga Hirosi-ni seetaa-o presento-si-ta
Yoko-NOM Hiroshi-DAT sweater-ACC present-do-PAST
Yoko gave a sweater to Hiroshi as a present.

At the point of parsing *seetaa-o*, a typical bottom-up parser would have no way to connect the three NPs into one structure because there are no heads available to license any structural commitments between the three NPs. One such strictly bottom-up head-driven parser is that of Pritchett (1991). In a discussion of Japanese NP sequences like the one in (91), Pritchett states that all of the NPs remain unattached until licensed by a head, in this case the verb (p. 262). Thus, his analysis of these structures is necessarily non-incremental.

Bader and Lasser (1994) provide evidence that a series of arguments in German is not parsed according to the predictions of Pritchett s head-driven parsing model, as discussed in ⁄ 1.2.2.1. Because the experiment is quite relevant to the question of whether processing is incremental in head-final structures, I will repeat the discussion here. In a self-paced reading experiment, Bader and Lasser investigated the head-final German structures in (92) and (93), exemplified in (94) in (95).

(92) [$_{CP2}$*dass* [$_{CP1}$ NP PP V1] V2]

(93) [$_{CP2}$*dass* NP [$_{CP1}$ PP V1] V2]

(94) [dass [sie$_{ACC}$ nach dem Ergebnis zu fragen] tats chlich erlaubt worden ist]
that she about the results to ask really permitted been is
that permission has indeed been given to ask her about the result

(95) [dass sie $_{NOM}$ [nach dem Ergebnis zu fragen] tats chlich erlaubt hat]
that she about the results to ask really permitted has
that she gave permission to ask about the result

The key to these sentences is that they are identical until the auxiliaries are encountered at the end of the sentence. Thus, they are ambiguous until just before the sentence terminates. Pritchett s model predicts that the pronoun *sie* is not analyzed until the first verb (*zu fragen*) is seen. At that point, the grammatical principles can be maximally satisfied by attaching *sie* as the object of the verb *zu fragen.* However, Bader and Lasser found that the preferred interpretation is for *sie* to be the subject of the verb *erlaubt*. They reason that this result is what would be expected if the parser is building structure incrementally. In particular, they assert that, in incremental processing, the pronoun *sie* will most naturally be interpreted a subject of some unseen verb, thereby making it unavailable for use as the object of *zu fragen*.

These results certainly provide evidence against the formulation of incremental parsing provided by Pritchett. While one might claim that the results can be explained by mechanisms other than incremental parsing (e.g. distributional characteristics of the verb *zu fragen* or a general preference to use *sie* as a subject), the results are certainly suggestive of incremental parsing. Several other pieces of converging evidence pointing towards incremental processing in head-final constructions are presented below.

In an effort to show that there is some syntactic analysis in Japanese before any licensers are encountered, Mazuka and Itoh (1995) reported an eye-tracking study that involved a series of three nominative NPs (at the beginning of a grammatical sentence). They contrasted this with a NOM ACC DAT series, and found that during the course of parsing the NPs (before any verbs are encountered), the all-nominative series is read significantly more slowly. They take this as evidence that some syntactic processing is performed immediately, since the two groups should be read identically if no processing at all takes place until a verb is encountered.[25]

Further arguments for the (strict) incrementality of Japanese parsing can be found in Inoue and Fodor (1995). One of their arguments is that the three NPs in (96) are preferentially understood as arguments to a single verb, resulting in a surprise effect at the monotransitive verb in (98) but not at the ditransitive verb in (97). They note that the surprise effect is not the same as normal garden path effects. The surprise is simply that readers are expecting a different type of verb, not that they have any difficulty understanding the verb.

(96) Bob-ga Mary-ni ringo-o
Bob-NOM Mary-DAT apple-ACC

(97) Bob-ga Mary-ni ringo-o ageta.
Bob-NOM Mary-DAT apple-ACC gave
Bob gave Mary the apple.

(98) Bob-ga Mary-ni [ringo-o tabeta] inu-o ageta.
Bob-NOM Mary-DAT apple-ACC ate dog-ACC gave
Bob gave Mary the dog that ate the apple.

[25] An alternative account (that does not assume incremental structure-building) would be that it is more difficult to remember/memorize a series of nominative NPs because they are all so similar. The NOM ACC DAT series might be easier because the NPs are easier to remember because of their distinctness.

Inoue and Fodor argue that this example provides evidence that this series of NPs is immediately analyzed as arguments of a single ditransitive verb, even before the verb is encountered in the input.

There are, of course, alternative interpretations. One possible interpretation of the surprise effect is that people are surprised that the verb does not use all of the available arguments. In particular, once the verb has been processed, it still cannot be attached into a single tree for the entire sentence. This alone might be enough to cause surprise effects. Additionally, the fact that a null operator needs to be built (as part of the relative clause) might also be a source of both surprise and increased syntactic processing. The German examples below in (103) and (106) could presumaby be used to build similar arguments, but would not be susceptible to the criticism that extra phonologically null structure be built in one and not the other (i.e. (106) does not require that relative clauses and the concomitant null operators be built.)

Support for incremental syntactic processing can also be found in the following German sentences:

(99) Die Frau glaubt, dass der Junge wegen seiner schlechten Noten sich selbst
the woman thinks that the boy because-of his bad grades him/her self
erschossen hat.
shot has
The woman thinks that the boy shot himself because of his bad grades.

(100) Der Mann glaubt, dass der Kollege wegen Geisteskrankheit ihn
The man thinks that the colleague$_{\text{MASC}}$ because-of mental-illness him
erschossen hat.
shot has
The man $_i$ thinks that the colleague$_j$ shot him$_i$ because he was mentally ill.

In (99), there are no morphological markings on the reflexive *sich selbst* to indicate whether the antecedent is masculine or feminine, and in (100) there are no morphological markings on the pronoun *ihn* to disambiguate between *Der Mann* and *der Kollege* as the antecedent. Despite the fact that the words themselves provide no hints about which NP should be the antecedent, native speakers report that they know immediately that *der*

Junge is the antecedent of *sich selbst* in (99), and they also know immediately that *der Kollege* cannot be the antecedent for *ihn* in (100). On the reasonable assumption that binding theory requires syntactic structure to operate (see footnote 1 on page 2), these sentences also provide evidence for incremental processing in head-final structures.[26] If structure is not built until a head that can license the different parts of these embedded clauses is encountered, there should be no structure available for the binding theory to operate over. The fact that the native speakers I have consulted have clear intuitions about binding facts before the verb is reached suggests that sufficient structure for the binding theory to be applied has been built before the verbs are reached.

3.2 Head Final Ambiguity

Given the evidence presented above suggesting that head-final structures are interpreted before the (licensing) heads have been seen, one might conclude that the parsing system commits to a particular analysis, and that only continuations consistent with that analysis can be easily processed. Many of these sequences are structurally ambiguous, and despite the fact that syntactic processing apparently takes place before the licensing heads are encountered, there is considerable flexibility in how these ambiguous strings of NPs can be completed and interpreted (i.e. building structure to allow a particular analysis does not preclude any of the other grammatical analyses). A theory that requires commitment to a particular analysis in advance of the disambiguating heads predicts that it should be relatively easy to garden-path speakers in head-final languages. However, as is shown below, it is not easy to cause parsing breakdown in such speakers.

This section presents several ambiguous sequences of NP arguments that are compatible with a number of different analyses, all of which can be easily parsed by the native speakers I have consulted. Consider first the simple case of a nominative NP followed by an accusative NP in a German embedded clause:

---

[26] This notion was also suggested in Bader 1994 as a diagnostic for incremental structuring in head-final languages.

(101) **NP_{NOM} NP_{ACC}** V_{TRANS}
dass er den Hund sah
that he $_{NOM}$ the dog_{ACC} saw
that he saw the dog

(102) **NP_{NOM} [NP_{ACC}** V_{TRANS}] V_{TRANS}
dass er den Hund zu f ttern vergass
that he $_{NOM}$ the dog_{ACC} to feed forgot
that he forgot to feed the dog

In this example, the two NPs can be interpreted as subject and object of a single transitive verb as in (101), but they are also compatible with a control structure, as seen in (102). According to my informants, neither of these continuations is difficult to process.

The examples in (103)-(106) show four possible continuations for a sequence of three NPs.[27]

(103) **NP_{NOM} NP_{ACC} NP_{DAT}** $V_{DITRANS}$
dass er den Hund dem Kind gab
that he $_{NOM}$ the dog_{ACC} the $child_{DAT}$ gave
that he gave the dog to the child

(104) **NP_{NOM} [NP_{ACC} NP_{DAT}** V_{DAT}] V_{TRANS}
...dass er den Hund dem Kind entfliehen sah
that he_{NOM} the dog_{ACC} the $child_{DAT}$ flee saw
that he saw the dog flee from the child

(105) **NP_{NOM} [NP_{ACC} NP_{DAT}** $V_{DITRANS}$] V_{TRANS}
dass er den Hund dem Kind gekauft zu haben bereut
that he the dog the child bought to have regrets
that he regrets having bought the dog for the child

(106) **NP_{NOM} NP_{ACC} [NP_{DAT}** P] V_{TRANS}
dass er den Hund dem Kind zuliebe gekauft hat
that he the dog the child to-love bought has
that the bought the dog to please the child

None of these continuations causes noticeable trouble for native speakers, despite the fact that they require the three NPs to be used in four different syntactic configurations.

[27] I thank Owen Rambow for help in identifying possible continuations of this ambiguity.

In all of the German examples presented above, the NPs have all been licensed by one contiguous string of heads. One way to account for the ease of all of these different analyses and to still assume that a single structure is built up incrementally, prior to the licensing heads, is to assume that all of the licensing heads are processed as if they were a single head. Under this assumption, one could conclude that all of these clauses are initially analyzed as ditransitives, with no need to change the structure for any of the continuations (c.f. the clause union analysis of Evers 1986). Thus, the need for flexible syntactic structure might be eliminated. However, the Korean examples in (107) and (108), like the Japanese examples in (97) and (98) above, show that this is not the case. In these examples, which are both parsed easily by native speakers, the licensing heads are broken up by the presence of NPs heading relative clauses.[28]

(107) $\mathbf{NP_{TOP}}$ $\mathbf{NP_{DAT}}$ $\mathbf{NP_{ACC}}$ $V_{DITRANS}$
na-nun John-eykey ku ai-lul tayliko ka-ss-ta
I-Top John-to the child-ACC to-take go-PST-DEC
I took the child to John.

(108) $\mathbf{NP_{TOP}}$ $\mathbf{NP_{DAT}}$ [$\mathbf{NP_{ACC}}$ V_{TRANS}] NP_{ACC} $V_{DITRANS}$
na-nun John-eykey ku ai-lul hayli-n salam-ul tayliko ka-ss-ta
I-Top John-to the child-ACC hit-REL person-ACC to-take go-PST-DEC
I took to John the person who hit the child

Summing up this section, the sentences in (101)-(108) show that any structure generated by the parser for head-final languages needs to be flexible enough to allow for multiple possible continuations without requiring extensive revisions. The major puzzle at this point is how to reconcile the fact that head-final structures appear to be processed incrementally with the fact that there is considerable freedom in the disambiguations that do not cause parsing breakdown. This is particularly puzzling in light of the most straightforward option for incremental structuring commitment to a single grammatical analysis. In the following section I show that structures can be built that allow for considerable flexibility in disambiguation, without the need for retraction of assumptions.

[28] These examples from Kisuk Lee (p.c.)

3.3 Building Flexible Structure

The basic method adopted here for dealing with the flexibility required for the examples in ⁄ 3.2 is to build heads that are underspecified for some features. This follows closely from the account presented in Chapter 2 for English. For example, predicted heads will be built that contain nothing more than case-assigning features, as in [Case: ACC, Left]. There is no general requirement that predicted heads contain categorial features, agreement features, or any other sort of feature. The only requirement for a predicted head is that it contain at least one feature capable of licensing the head that it is attached to.

Note that this featural underspecification differs from the node underspecification used in D-theory. D-theoretic work like that of Weinberg (1993) involves the use of domination relations to specify syntactic structure, rather than the more common immediate domination. The use of domination rather than immediate domination means that extra heads can be added into a tree without the need to retract any domination statements (i.e. with monotonically increasing information). This is true because the domination relation between two nodes continues to hold if extra nodes are inserted between them. If two nodes are related by an immediate domination relation, the insertion of extra nodes between the two means that the higher node no longer immediately dominates the lower node. Thus, the immediate domination relation between the two would need to be retracted (see discussion in ⁄ 1.2.3.2 and ⁄ 5.4.1.2 for more details). The structures built by SPARSE contain fully-specified immediate domination links, which means that some structure (i.e. domination relations) must be retracted whenever a syntactic node is added between two existing nodes. Because structure-building in SPARSE is not required to be strictly monotonic, extra syntactic nodes can added whenever circumstances require it.

To get an idea of how SPARSE builds structure in head-final languages, consider (109) and (110) (repeated from (101) and (102) above).

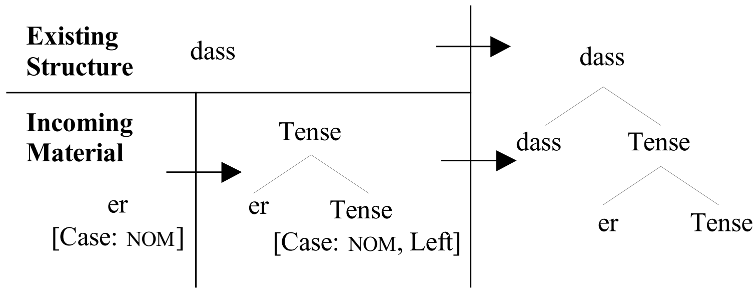
(109) **NP$_{NOM}$ NP$_{ACC}$** V$_{TRANS}$
dass er den Hund sah
that he $_{NOM}$ the dog$_{ACC}$ saw
that he saw the dog , e.g. *I knew that he saw the dog*.

(110) **NP$_{NOM}$ [NP$_{ACC}$** V$_{TRANS}$] V$_{TRANS}$
dass er den Hund zu f ttern vergass
that he $_{NOM}$ the dog$_{ACC}$ to feed forgot
that he forgot to feed the dog

Both of these sentences are easy to parse. Below is an explanation of the parsing process common to the two clauses (i.e. up to the verb(s)), followed by an explanation of how the verbal material in each of the clauses is processed.

For the purposes of this example, I will assume that the complementizer *dass* is the first word of the sentence (i.e. it does not need to attach to anything else). Of course, in most real examples *dass* will already be attached as part of an existing sentence, but this does not affect the issues discussed here. When the nominative pronoun *er* he is encountered, it cannot be directly attached to *dass*, because *dass* does not assign the nominative case that *er* requires. Because *er* cannot be attached into the main tree, the distinguished feature for *er*, nominative case, is used as the basis for a licensing head. When the lexicon is searched for heads that can assign nominative case, the results show that all heads that assign nominative case also have tense specifications (along with the other features that go along with tense heads like agreement features). Accordingly, a Tense head is posited as the licenser for the pronoun, and the entire pronoun-tense complex is attached as the complement of the complementizer *dass* (which selects for a head with a tense feature).

(111)



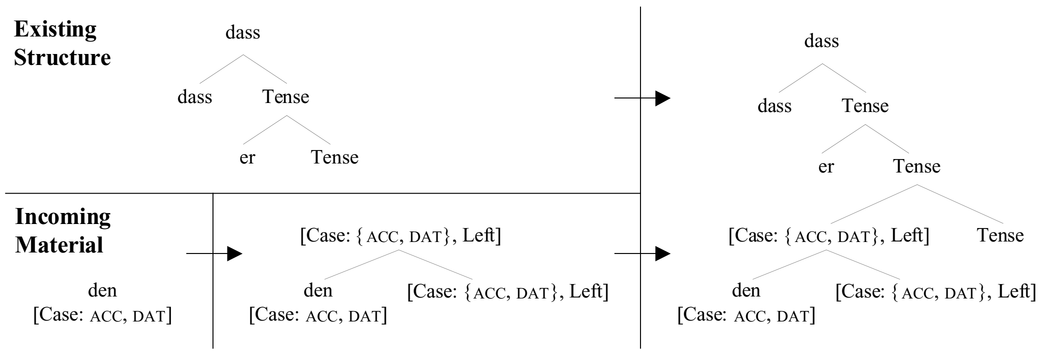
Because the entire lexicon is searched for heads that can license nominative NPs and only one type of head is returned, it can be guaranteed that a Tense head will license *er* in the tree. The basic idea behind building predicted heads is that all features that are in the intersection of the heads that meet the search criteria will be built into the predicted head, but no other features will be present. In this case, because Tense heads are the only heads that assign nominative case in German, only tense heads will be returned from the search, and the intersection of tense heads is a tense head.

When *den* the is encountered, it again cannot be directly attached to the existing structure. The determiner itself is ambiguous between masculine-accusative and plural-dative. As a result, the case feature at the top of its stack is also ambiguous between accusative and dative. The lexicon is searched for heads that can license either accusative or dative NPs. This search returns a variety of heads, including dative and accusative verbs, as well as both dative and accusative postpositions. The only thing common to all of these heads is the fact that they can assign case. Accordingly, a case-assigning head is built and attached to the right of the determiner. The value of the case-assigning feature is initially set to a variable (since the lexicon search/intersection returned a completely underspecified value), when the case-assigner is attached to the determiner, the case-assigning features enters into a checking relation with the determiner. Accordingly, the value of the case-assigning feature is set to {ACC, DAT}, since those are the only values which are allowed by *den*. Even with a case-assigning head attached to the determiner *den*, there is still no direct way to attach *den* into the existing tree, since nothing in the tree selects for a case-assigner.

A search is next conducted for an attachment site for the *den* + Case-assigner constituent. Because the case-assigner (the head of the constituent) is a predicted head that is underspecified for most of its features, the search for an attachment site requires something more than the usual search procedure. The case assigner only contains the information that it assigns either accusative or dative case, and a normal search of the lexicon will come up with no heads that can license it. The fact that the search returns a null result is related to the fact that no heads select for a complement that necessarily contains a case-assigning feature. Instead, heads select for other properties of their complements (e.g. Tense or [Cat: V]). If the case-assigning head were fully specified for all features (e.g. if it had a Category feature), the search would return a head that could license it. Because predicted heads are frequently underspecified for the features that other heads select for, an extra step needs to be added to the parsing algorithm.

The extra step that will be added to the algorithm specifies that the search for an attachment site for a predicted head involves using all heads that are compatible with the predicted head. In this case, the extra step means that the search for an attachment for the {ACC, DAT} case-assigner will actually involve a search for possible attachments of all of the different accusative and dative case-assigning heads. As noted above, these heads include accusative and dative verbs, as well as accusative and dative postpositions. While there are no possible attachment sites in the existing tree for postpositions (due to the fact that nothing in the tree selects for a PP or allows PP adjunction), both of the verb types (dative and accusative) could be attached as the complement of the Tense head. Accordingly, the case-assigner is attached as the complement of the tense head. Note that the case-assigner is not further specified to be a verb, as there is no guarantee that the case-assigner will ultimately take the form of a verb (i.e. the bottom-up information from the determiner does not provide enough information to decide the question). The only thing that is certain about the licenser is that it will assign either accusative or dative case to the determiner, thus that is the only information that will be included in the predicted head.[29]

(112)



This method of predicting underspecified heads bottom-up from the incoming material is very similar to the method used in Konieczny s (1996) SOUL parser to build licensing heads (e.g. Konieczny uses a Det head to predict an NP).[30] In the HPSG analysis used by Konieczny, a verb is predicted on the basis of the subject, and it is not necessary to predict extra heads to license any complements of the verb. Thus, Konieczny does not use the prediction mechanism to predict new licensing heads for complements. Instead, the verb that is predicted on the basis of the subject is underspecified for its complements. When the complements are encountered, they allow for a further specification of the complement features of the predicted verb. Thus, Konieczny predicts that the nominative and accusative NPs will be licensed by the same verb. However, we will see below that by adding the features of all of the complements into a single head, SOUL is unable to easily account for many of the temporary ambiguities in head-final languages.

Figure 3 below shows the parsing algorithm with the extra step needed for parsing these types of head-final constructions. Changes to the algorithm are italicized.

[29] Because structure is built strictly on the basis of bottom-up information, the requirement that sisters enter into a licensing relation must be relaxed slightly to a requirement that sisters be able to enter into a licensing relation.

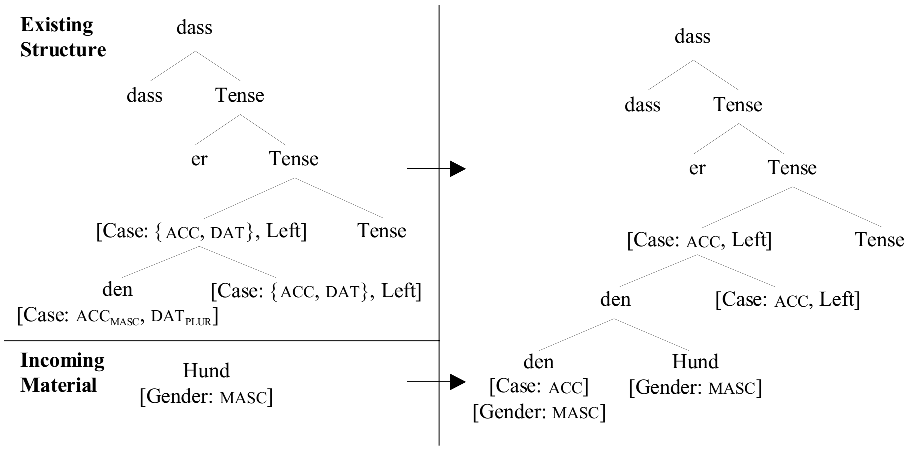
[30] Konieczny assumes an NP analysis rather than the DP analysis used in this dissertation.

1. Search unchecked features on the right edge of the existing structure for an argument attachment for the incoming material.
2. Search the right edge of the existing structure for an adjunct attachment of the new material.
3. Build a new licenser for the new material:
   3.1. Search the lexicon for all possible licensers of the new material.
   A possible licenser is either a head with appropriate left-pointing features or a null head with appropriate right-pointing features.
   3.2. Attach to the new material the intersection of all heads returned by the lexicon search.
      - *If the intersection is null and the new material is headed by a predicted licenser, search for an argument attachment using all heads compatible with the head of the new constituent. If successful, make the attachment using the head of the new constituent (not the subsuming head that licensed the attachment)*.
      - If the intersection is null *and the new material is not headed by a predicted licenser,* no new licensing heads can be predicted continue on to 4, otherwise return to step 1 with the just-built constituent.
4. Search all features (checked and unchecked) on the right edge of the existing structure for an argument attachment for the incoming material.
   4.1. If attachment is found, remove existing element from the tree, attach it to the new material, and start over with newly expanded constituent at step 1.

Figure 3: SPARSE Parsing Algorithm (Version 3 of 4)

When the noun *Hund* dog is encountered, it can be immediately attached as the complement of the determiner *den*. The masculine gender and singular number of the noun allows for further specification of the gender, number and case of the determiner, turning it into a masculine, singular, accusative determiner. This revision of the features on the determiner also allows a further specification of the features on the case-assigner, with the result that the case-assigner is specified for only accusative case.

(113)



Once this structure has been built, the licensing head(s) are processed. I will first discuss the processing of transitive verb *sah* saw in (109), repeated as (114), and will then show the processing of the two verbs in (110).

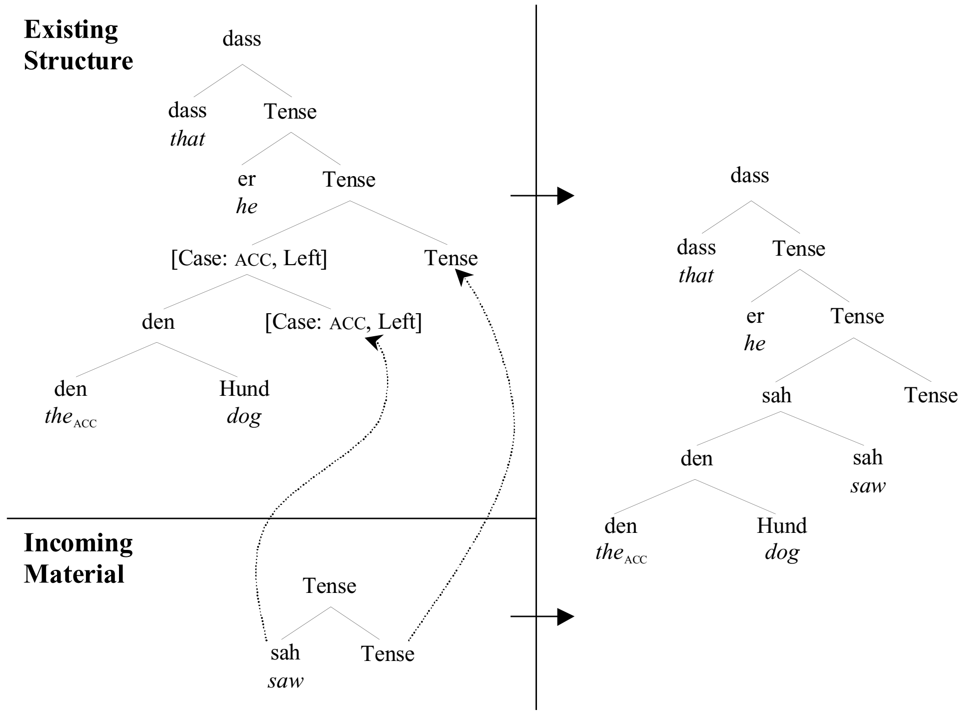
(114) $\mathbf{NP}_{\mathbf{NOM}}$ $\mathbf{NP}_{\mathbf{ACC}}$ $\mathrm{V}_{\mathrm{TRANS}}$
dass er den Hund sah
that he $_{\mathrm{NOM}}$ the $\mathrm{dog}_{\mathrm{ACC}}$ saw
that he saw the dog

I assume that the lexicon/morphology returns a complex structure for a finite verb like *sah*. In this case, I assume that the structure is that of a verb plus a tense head. In the implemented version of the parser this is done by storing the V+T complex as the lexical entry for the verb. In the human parser, the morphology presumably provides this structure to the syntactic parser. When a multi-headed structure for a new word is provided to the parser, the parser works by trying to find an attachment for the lowest head of the incoming item (e.g. the verb in a T-V complex lexical item). Because the lowest head (the verbal head) in the incoming word will (at least sometimes) subsume a predicted head (e.g. the case-assigner in (113)), the search must start at the lowest head in the incoming word (this is the case for all complex incoming material). If a higher head (e.g. the Tense head of the T-V complex) in the incoming word were tried first, the search would fail at the first predicted head (the case-assigner), because the search would not be able to proceed post the first predicted head to discover that a higher head (the T head) can be subsumed by the higher head in the incoming word.

Recall that predicted heads must eventually be instantiated by a real head (either a head in the input or a null head from the lexicon). Predicted heads are instantiated when a real head is found that is compatible with all of the features of the predicted head. A predicted head is compatible with a real head if the predicted head can be subsumed by the real head. In this context, subsumption means that every feature of the predicted head is also in the real head, and the intersection of the feature values for each feature in the predicted head is non-null. For example, there is a predicted head containing the feature [Case: ACC, Left] in the final tree in (113). This predicted case-assigning head can be subsumed by verb *sah,* because *sah* also contains an accusative-assigning feature to its left. There is no restriction against the real head having features that are not present on the predicted head.

Once it has been determined that the verb is compatible with a head in the existing tree (i.e. the case-assigner can be subsumed by *sah*), the rest of the heads in the incoming item are checked to see if they are also compatible with predicted heads in the existing tree. The features of the two tense heads are compatible, so the incoming heads can be incorporated into the existing tree. This process can be seen in (115).

(115)



Example (110), repeated here as (116), shows another possible continuation for the NOM ACC sequence that was analyzed in (113).

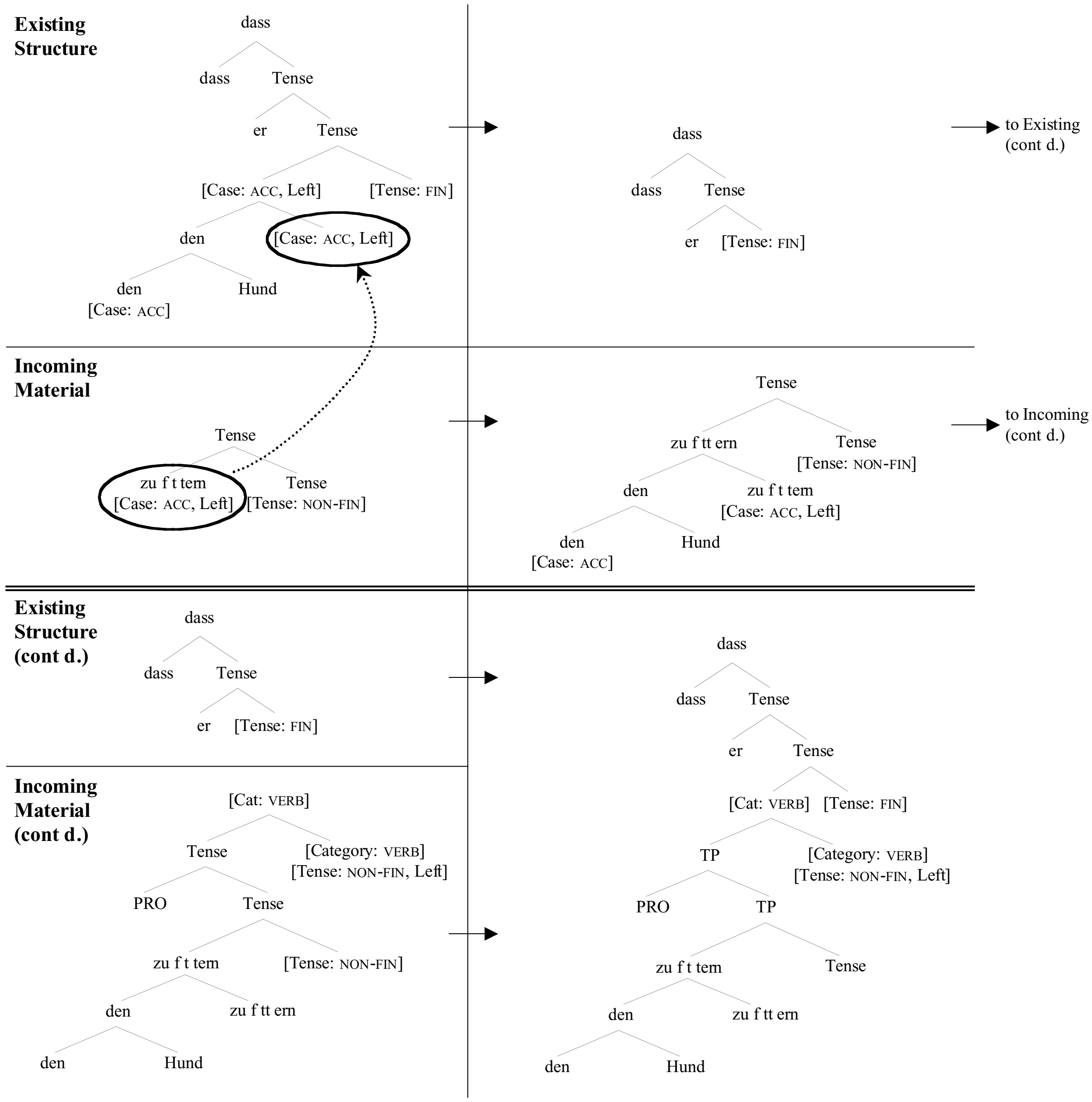
(116) **NP_{NOM} [NP_{ACC}** V_{TRANS}] V_{TRANS}
dass er den Hund zu f ttern vergass
that he $_{NOM}$ the dog_{ACC} to feed forgot
that he forgot to feed the dog

Since there is no difference in the input strings until the verbal material is reached, the NPs are processed the same as in (113); this discussion starts immediately after the two NPs have been parsed. I simplify slightly and assume that *zu f ttern* (to feed) is processed as a single word, and is available to the parser as a non-finite tense head with a verb complement (see tree in (117)). There is no way for the incoming verb-T structure to attach directly into the tree, since there is nothing in the existing structure that selects for this type of material. Likewise, the new heads cannot subsumed by the predicted heads in the existing tree, because the new tense head is non-finite, while the tense head in the existing structure is finite. The solution to this problem is to allow a portion of the existing tree to be removed if a subsumption relation can be found for a subset of the heads in the incoming material. In this case, a subsumption relation can be found between the predicted accusative assigner and the incoming verb. Thus, the section of the existing tree headed by the accusative case-assigner is removed and the case-assigner is subsumed by the verb. Removal of this portion of the existing tree differs only in its trigger from the stealing operation that is needed for reanalysis of NPs in English NP-S ambiguities.

The resulting structure, with the accusative NP, the verb and the tense head (shown in part 2 of (117), is then parsed as if it were all part of the incoming word. In this case, no more structure can be built above the tense head unless the external theta role is assigned to the left. A search is conducted for null elements that can take theta-roles (but don t need case). This search returns PRO. Accordingly, a PRO head is attached to the tense head. Once this has been done, a head to license the non-finite IP can be built. As noted earlier, the search is limited to heads selecting Tense on the left or non-overt heads selecting Tense on the right in order to preserve the linear order of the sentence. The only head returned by the search is a verb that selects for non-finite Tense, which is then attached to the Tense head. A search for an attachment site for the predicted verb is initiated, with the result that the predicted verb can be attached as the complement of the tense head in the existing tree. This entire process is illustrated in (117).[31]

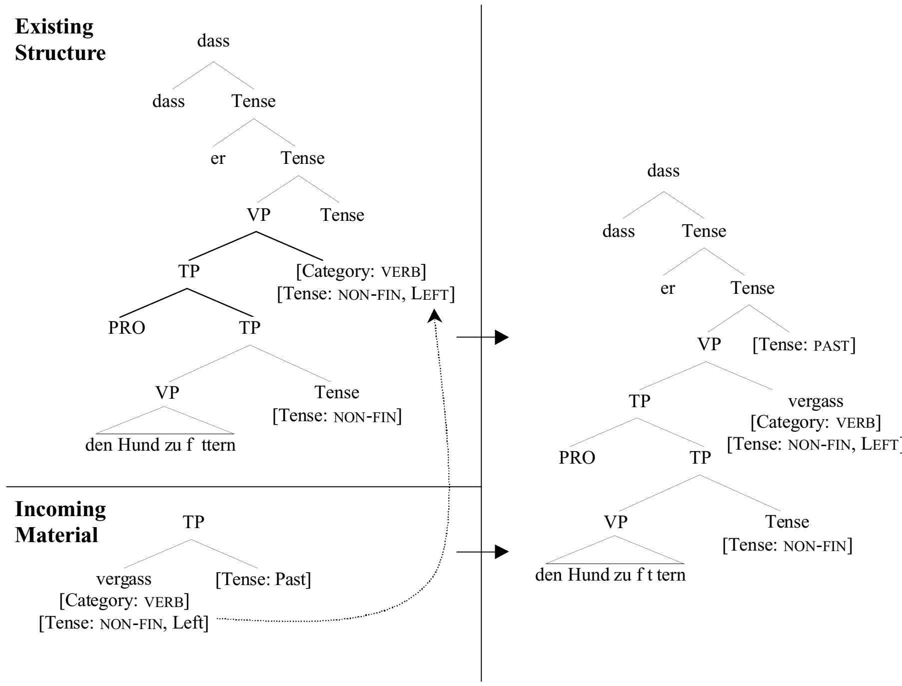
[31] The end result of this stealing is very similar to the tree-lowering used in monotonic accounts of structure-building (e.g. Weinberg 1993, Sturt and Crocker 1996). However, the mechanisms differ, as do the predictions. In particular, the monotonicity accounts predict that sub-trees can only be lowered, while SPARSE predicts that it should be possible to raise as well (given the proper syntactic conditions).

(117)



At this point, the only predicted heads in the existing trees are the matrix verb and the matrix tense head. When the verb *vergass* is encountered, its lexical entry is a V-T combination, which can be easily merged with the V and T heads present in the tree. This last bit of processing is shown in (118) below.

(118)



The parsing algorithm requires a number of changes in order to split the tree and re-assemble it as described above for (116). Figure 4 shows the revised parsing algorithm. The additions required to split and reassemble the tree are shown in italics.

1. Search unchecked features on the right edge of the existing structure for an argument attachment for the incoming material.

- *If a subsumption relation is found between all heads in the incoming item and predicted heads in the existing structure, integrate the entire new item into the existing tree.*
- *If subsumption is found between some (but not all) heads in incoming item and heads in the existing tree, remove from the existing tree the portion that is compatible with the heads in the incoming item. Integrate the new item into the just-removed structure. Return to step 1 with the just-integrated new item.*

2. Search the right edge of the existing structure for an adjunct attachment of the new material.
3. Build a new licenser for the new material:

    3.1. If requirements must be satisfied on the new item before structure can be built above it, build the minimum structure necessary to satisfy the requirements and continue to 3.2.

    3.2. Search the lexicon for all possible licensers of the new material.
    A possible licenser is either a head with appropriate left-pointing features or a null head with appropriate right-pointing features.

    3.3. Attach to the new material the intersection of all heads returned by the lexicon search.

    - If the intersection is null and the new material is headed by a predicted licenser, search for an argument attachment using all heads compatible with the head of the new constituent. If successful, make the attachment using the head of the new constituent (not the subsuming head that licensed the attachment).
    - If the intersection is null and the new material is not headed by a predicted licenser, no new licensing heads can be predicted continue on to 4, otherwise return to step 1 with the just-built constituent.

4. Search all features (checked and unchecked) on the right edge of the existing structure for an argument attachment for the incoming material.

    4.1. If an attachment is found, remove existing element from the tree, attach it to the new material, and start over with newly-expanded constituent at step 1.

Figure 4: SPARSE Parsing Algorithm (Version 4 of 4)

3.4 When to Build Predicted Material

An interesting question about predicted heads is the question of when or on what grounds they should be posited. Based on what has been presented so far, the answer is that predicted heads are only built when they are needed to allow for an incremental parse. However, Frazier (1987) suggests that there might be other reasons to posit predicted structure. Consider the Dutch sentence in (119), where the PP can either be attached as an adjunct to the NP, or it can be an argument of the upcoming verb.

(119) dat het meisje van Holland houdt/glimlachte
that the girl from/of Holland likes/smiled
..that the girl likes Holland / that the girl from Holland smiled

If the sentence is continued with *houdt* likes , the PP *van Holland* must be interpreted as an argument of the verb, while if it continues with *glimlachte* smiled, the PP can be interpreted as an adjunct modifying the NP *het meisje* the girl. In a self-paced reading study, Frazier tested one sentence similar to (119) as a filler in an experiment testing a different structure, and found that the analysis involving an adjunct attachment (with the intransitive verb *glimlachte*) was more difficult than the analysis as an argument of an upcoming verb. In the example in (119), this would mean that the structure in which the PP *van Holland* is attached as an argument of the upcoming verb should be easier to process than the structure in which the PP must be interpreted as an adjunct modifier of the NP *het meisje*. From this, Frazier predicts that the sentence ending in *houdt* should be easier than the sentence ending in *glimlachte,* because the PP would be initially attached as an argument, and then would need to be reanalyzed to an adjunct position when *glimlachte* is encountered. Frazier takes this as evidence that argument attachments are preferred over adjunct attachments, even in advance of the theta-assigner for the argument. Unfortunately, this structure was not systematically investigated in Frazier (1987).

Contrasting with Frazier (1987), Konieczny, et al. (1997) present evidence from an eye-tracking experiment showing that subjects read attributive PPs (i.e. NP-adjuncts as in

(120)) more quickly than instrumental PPs (i.e. attached to the VP as in (121)) in head-final German constructions. At the PP itself, the only thing that distinguishes (120) and (121) is the pragmatic bias of the PP towards the attributive (NP-adjunct) or instrumental (verbal argument) reading.

(120) Ich habe geh rt, dass Marion das Pferd mit dem weissen Fleck erblickte
I have heard that Marion the horse with the white patch saw
I heard that Marion saw the horse with the white patch.

(121) Ich habe geh rt, dass Marion das Pferd mit dem neuen Fernglas erblickte
I have heard that Marion the horse with the new binoculars saw
I heard that Marion saw the horse with/using the new binoculars

Konieczny, et al. reason that the slowdown in the argument-biased PPs in verb-final constructions is a result of the fact that all PPs are initially attached as NP-adjunct attachments, and the attachment must be changed when it becomes clear from the content of the PP that it should really be attached as an argument of the upcoming verb. Konieczny, et al. attribute their findings to their Head Attachment Principle, which states that the parser prefers to attach an item to a phrasal unit whose lexical head has already been reached. The (statistically reliable) results of Konieczny, et al. contradict the suggestion presented by Frazier (1987) that PPs are preferentially attached as arguments to upcoming verbs.

Because the experimental data from Konieczny, et al. (1997) was obtained from a rigorous, statistically-validated design, I have chosen to have SPARSE build its structure by attaching to existing material before positing predicted heads. Within SPARSE, this behavior is achieved by the fact that step 2 (adjunct attachment) precedes step 3 (build predicted licensing heads) in the parsing algorithm. As a result, all attachments to existing material are attempted before any new heads are predicted. With the simple step of ordering all direct attachments before the process of building predicted heads, all of the predictions made by the Head Attachment Principle can be made in this model as well.

3.5 Predictions

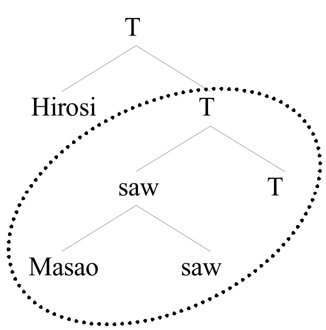
One type of reanalysis that should cause difficulty (but not breakdown) is the one that is required for Japanese sentences like (123) and (124) (from Mazuka and Itoh 1995). According to Mazuka and Itoh, sentences (122)- (124) can all be parsed without conscious difficulty. Immediately before *otoko-o* is encountered, (123) and (124) presumably have the structure of a simple SOV clause (as in (122), tree in (125)).

(122) Hirosi-ga Masao-o mita
Hirosi-NOM Masao-ACC saw
Hirosi saw Masao.

(123) Hirosi-ga [ﬂ$_i$ Masao-o mita] otoko-o $_i$
Hirosi-NOM Masao-ACC saw man-ACC
Hirosi did something to the man who saw Masao.

(124) Hirosi-ga [ﬂ$_j$ [ﬂ$_i$ Masao-o mita] otoko-o$_i$ yobidasita] onna$_j$
Hirosi-NOM Masao-ACC saw man-ACC called woman
Hirosi did something to the woman who called the man who saw Masao.

(125)



However, when *otoko-o* is encountered in (123), it can only be interpreted as the head of a relative clause. In order to build a relative clause, the entire VP-T constituent (the traditional T , circled in (125)) must be reanalyzed from a position inside the matrix clause to a location inside a relative clause. The only piece of the original tree that is not taken is the NP *Hirosi.* Thus, this operation effectively separates the subject from the rest of the clause. The T rather than the VP node must be reanalyzed (removed from the existing tree), because tense information is carried on the verb itself, and the tense head and the VP can not be separated.[32]

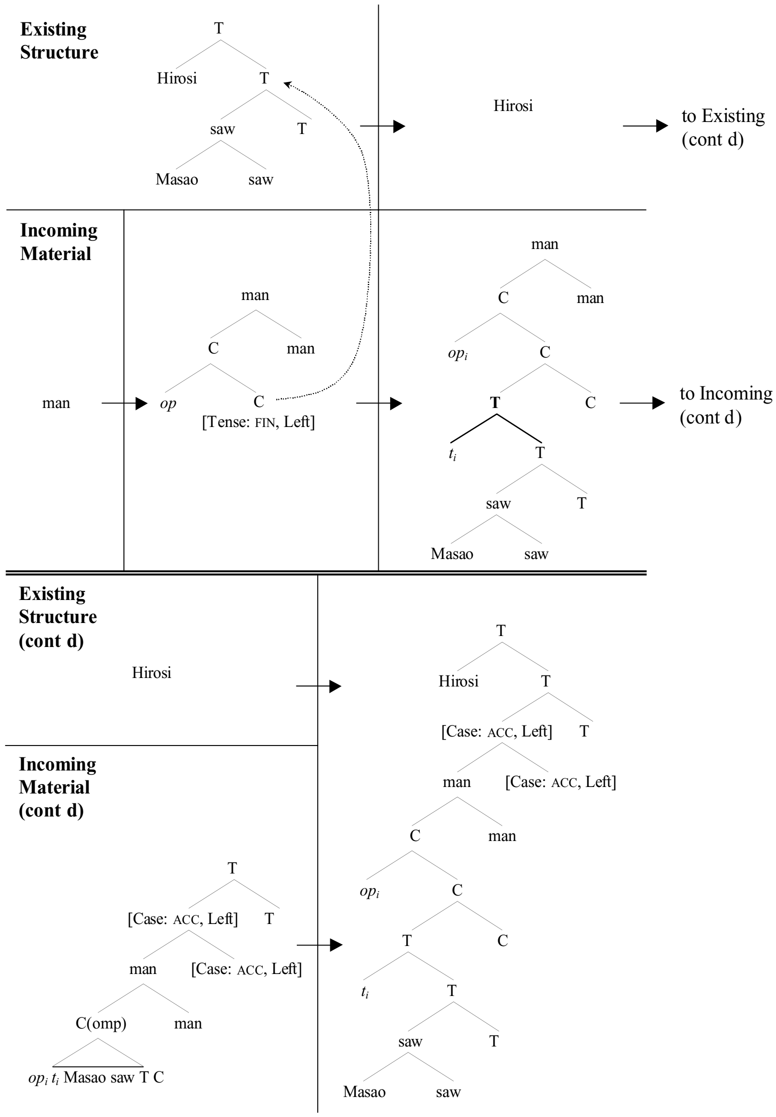
One complication in this example is the fact that the extra structure necessary for a relative clause must also be posited. This extra structure is built using the normal structure-building mechanisms. When the NP *otoko-o* is encountered, it cannot be attached directly to the existing tree. A search for predicted heads is therefore initiated. One of the heads returned by the search is a null C head for a relative clause (along with the associated operator, which I assume is stored with the CP that selects for an NP) that adjoins to the left of an NP. Although relative clauses are adjuncts, the C head of a relative clause selects for an NP, so it will be returned by the search for elements that select for NPs. Once this has been built, the I node from the existing tree can be taken as the complement of the CP. As a consequence of the operator in the Spec position of the relative clause C head, a trace is inserted as the subject of *saw*; a full discussion of how traces of movement are posited can be found in Chapter 4 below. Once the accusative relative clause has been assembled, it is attached to the existing tree (which now only consists of Hirosi-ga). Then, following the standard procedures for building predicted heads, the relative clause forces the construction of the accusative case assigner and tense head necessary to connect an accusative NP and a nominative NP. The process of building the relative clause is illustrated in (126) below.

[32] Consider the following sentence:

(v) Hirosi-ga [fl$_i$ Masao-o mita] otoko-o$_i$ sitte-iru
Hirosi-NOM Masao-ACC saw man-ACC knows
Hirosi knows the man who saw Masao.

If the past tense T head stayed with the subject, then it should not be possible to have a present tense verb for that matrix subject after reanalysis.

(126)



In essence, what looks superficially like the removal of a node from the left of the tree (e.g. *Hirosi* in (126)) can also be modeled as the removal of the entire right edge of the tree from the existing structure.

As can be seen from (124), it is also possible to perform this process of removing the subject more than once in a single sentence. At the point after the second verb *yobidasita* has been processed, the structure of the sentence is like final structure for (123), with *Hirosi-ga* serving as subject of *yobidasita*. However, when *onna* signals the presence of another relative clause, the entire process is triggered again and everything but *Hirosi* is reanalyzed into a relative clause.

While removal of one head from the left edge of the tree is possible (by actually removing everything else to its right), removal of more than one head from the left edge of the tree is beyond the capability of the parser. Consider the Japanese sentence in (127) (from Mazuka and Itoh 1995), where the only possible interpretation causes conscious, difficult reanalysis.

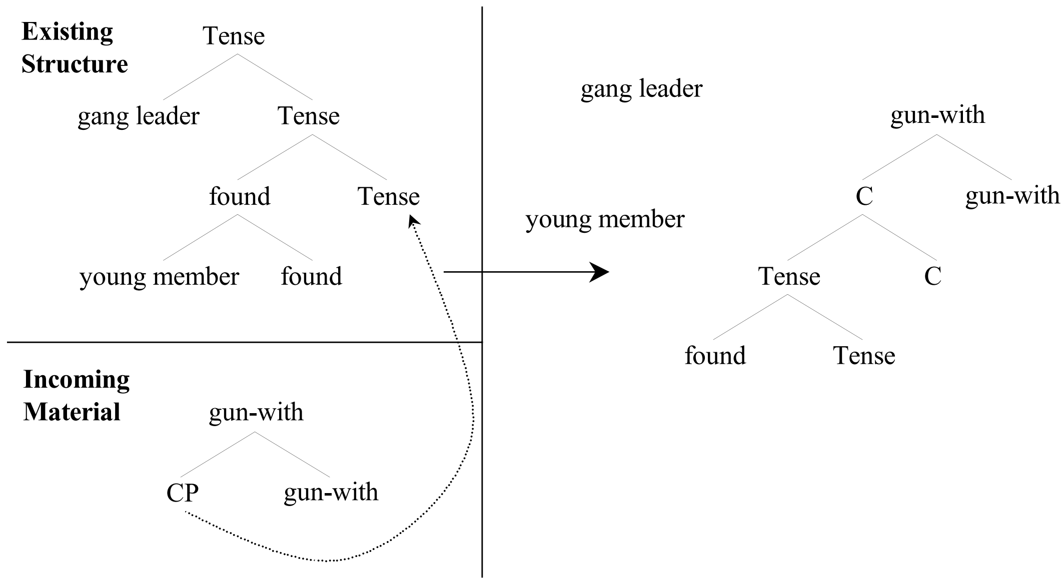
(127) Yakuza-no kanbu-ga wakai kobun-o sagasi-dasita kenzyuu-de
gang-GEN leader-NOM young member-ACC found gun-with
utikorosite simatta
shot to death
The gang leader shot the young member to death with the gun he found.

In this sentence, the gang leader and the young member are initially analyzed as the subject and object of found . However, when *kenzyuu-de* with a gun is encountered, both previous NPs must be removed from the verb *found* and its associated Tense head. Native speakers report conscious difficulty when they read *kenzyuu-de*, which I take to indicate that the reanalysis is beyond the ability of the automatic parser. Thus, it appears that removal of nodes from the right-hand side of the tree is limited to just one head.

To see how this limit is predicted in SPARSE, consider the tree in (128).[33] If a head that the Tense head (and its associated verb) as its complement to the exclusion of *gang leader* and *young member*, the only possible way to achieve this is to strand both *gang leader* and *young member* as separate trees. This has the effect of leaving too many left over pieces for the parser to be able to successfully put back together.

[33] The tree in (128) contains a very much simplified structure for the relative clause. The discussion also ignores extra complications, such as how just the Tense and verb heads could ever be removed without the subject and object NPs that are attached to them.

(128)



Throughout the discussion of the parser, we have seen that the parser can work with tree for the existing tree and the incoming word, but is not capable of dealing with any other constituents. Thus, this example is correctly predicted to cause parsing breakdown.

The previous example showed that only one item can be removed from the left edge of the tree (by actually removing the entire right edge of the tree). Recall that in order to reanalyze the entire right edge of the tree, a non-maximal projection had to be removed from the existing tree. The limit on removing one element from the left side of the tree actually reflects a limit on which non-maximal projections can be reanalyzed. Specifically, the only non-maximal projection that can be reanalyzed is the one immediately below the root of the tree. As was discussed in relation to the tree in (128), removal of a non-maximal projection anywhere but at the root of the tree leaves too many disconnected pieces of structure, causing parsing breakdown. For this reason, I will assume that the search of the tree for reanalysis sites only considers maximal projections except for the projection immediately below the root of the tree.

A number of other explanations have been proposed in the literature for difference in difficulty between (122)-(124) and (127). Mazuka and Itoh (1995) claim that the difference in difficulty is due to the fact that two NPs are reanalyzed in (127) to a position associated with a different verb, while only one NP is reanalyzed in (122)-(124). In their theory, an individual reanalysis has some cost, but the cost is not high enough to cause conscious difficulty. However, the costs of multiple reanalyses, as in (127), accumulate to cause conscious difficulty. Thus, under their theory, it is irrelevant whether reanalysis is taking a piece from the left or the right side of the existing tree. Their theory therefore predicts that a sentence requiring multiple reanalyses on the right edge of the tree should be just as difficult as a reanalysis involving multiple reanalyses on the left edge of the tree. If this extension of Mazuka and Itoh s theory is correct, it predicts that a German sentence like (130) should also be a conscious garden path, since it requires that both *dem Kind* the child and *das Fahrrad* the bicycle be reanalyzed from objects of *verspeche* promise to arguments of *zu geben* to give.

(129) Ich [verspreche dem Kind das Fahrrad].
I promise the child$_{DAT}$ the bicycle$_{ACC}$
I promise the child the bicycle.

(130) Ich verspreche [dem Kind das Fahrrad zu geben].
I promise the child$_{DAT}$ the bicycle$_{ACC}$ to give
I promise to give the child the bicycle.

However, according to the native speakers I have consulted, this sentence does not cause conscious reanalysis, and is therefore not beyond the limits of the automatic parser.

A number of explanations of the difference between (127) and (122)-(124) have also been made within the D-theory literature (Gorrell 1995, Sturt and Crocker 1996, Weinberg 1992). Gorrell and Weinberg show that the change necessary for (127) is not monotonic in their systems, while Sturt and Crocker argue show that the reanalysis necessary for (127) is possible, and they give an example of the same sort of ambiguity that they claim does not cause conscious reanalysis.

It is difficult to determine whether any of these theories would predict parsing breakdown for the sentence in (130) because it is not clear exactly how the verb *verspreche* would be analyzed (i.e. are traces of it found in the heads that license the two object NPs?). If the verb (or features of the verb) is found in the heads licensing the two objects, the theories of Gorrell and Sturt and Crocker predict that this sentence should cause breakdown. They predict this because of additional restrictions on addition of structure. In both theories, it would be impossible to move the first object , the dative NP *dem Kind* from a position preceding a projection of the verb (the head licensing the accusative NP *das Fahrrad*) to a position after the verbal projection. The theory proposed in Weinberg does not contain restrictions limiting this sort of change, and therefore does not predict that this sentence should cause parsing breakdown.

Chapter 4

LEFTWARD MOVEMENT

4.1 Introduction

So far, this thesis has been concerned with the processing of sentences with canonical word order. This chapter discusses how wh-movement permutes canonical word order and how wh-movement is handled within SPARSE. This introductory section discusses basic movement facts in English, including some of the major constraints on movement. ⁄4.2 discusses how movement is handled within the SPARSE framework, and ⁄4.3 shows how this approach captures a series of constraints on movement that are discussed below. A discussion of predictions made by the theory is found in ⁄ 4.4, and a discussion of how the theory is related to the grammar completes the chapter in ⁄ 4.5.

The discussion of wh-movement will begin with the very simple example below, where the location of the object changes from after the verb in (131) to before it in (132).

(131) Toto chased her.

(132) Who$_i$ did Toto chase fl$_i$?

This example illustrates a general fact that when questioned, most NPs can be realized at the beginning of a sentence rather than in their canonical locations. Wh-movement can be seen in English with pronominal NPs (e.g. (132)), full NPs (e.g. (133)), and PPs (e.g. (134)).

(133) Which witch$_i$ did Dorothy s house land on fl$_i$?

(134) To whose residence$_i$ did the monkeys take Dorothy fl $_i$?

Wh-movement is not limited to single clauses. As can be seen in (135) and (136), wh-phrases can move across multiple clause boundaries.

(135) What$_i$ did the scarecrow say [$_S$ he wanted fl $_i$]?

(136) What$_i$ did Dorothy think [$_S$ the scarecrow said [$_S$ he wanted fl $_i$]]?

While wh-elements can be moved fairly liberally in languages that allow wh-movement, there are a number of restrictions on when this movement can occur, the so-called island constraints. Four of the best-known constraints, first noticed by Ross (1967), will be discussed here; other related facts will be discussed below. The basic idea behind island constraints is that certain constituents do not allow wh-elements to move out of them they form seemingly impervious barriers to extraction.

One of the islands that Ross discussed is the wh-island. Wh-islands are embedded clauses that begin with wh-elements that block extraction of wh-elements from inside the constituent, such as indirect questions and relative clauses. Examples can be seen in (137) and (138).

(137) Dorothy wondered [how the wizard could make the ferocious noises].

(138) Dorothy met a scarecrow [who was missing a brain].

Ross noticed that it is generally not possible to question out of wh-islands. Thus, sentences like (139), where an element inside the wh-island is questioned, are ungrammatical.

(139) *How$_i$ did Dorothy wonder [which noises$_j$ the wizard could make fl$_j$ fl $_i$]?

The example above shows that adjuncts cannot be extracted from wh-islands. The example in (140) below, however, shows that extraction of arguments is somewhat better (Huang 1982, Lasnik and Saito 1984 and references therein).

(140) ??Which noises$_i$ did Dorothy wonder [how$_j$ the wizard could make fl$_i$ fl $_j$]?

Because of the fact that arguments can be extracted from them (though at some cost), wh-islands will be considered to be weak islands (Huang 1982). A number of other types of islands pattern with wh-islands as being more susceptible to argument extraction than adjunct extraction. Among the other weak islands are factive islands, negative islands, and definiteness islands. Examples of these types of islands can be seen in (141)-(146) below (examples from Szabolsci and Zwarts 1991 and Melvold 1991).

(141) ?Which man$_i$ did you deny [that John fired fl$_i$]?

(142) *Why$_i$ did you deny [that John was fired fl$_i$]?

(143) ?Which man$_i$ don t you think [that John fired fl$_i$]?

(144) *Why$_i$ don t you think [that John was fired fl $_i$]?

(145) ??Who$_i$ did John find [my picture of fl $_i$]?

(146) *Where did John find [my picture of Sarah fl$_i$]?

Another type of constituent that forms an island to extraction is the complex NP, as in *a report that the wizard said something nasty* in (147). The ungrammaticality of (148) is due to the fact that *what* is moved out of the complex NP headed by *a report*. This type of island is considered to be a strong island, because neither arguments nor adjuncts can be extracted.

(147) Dorothy heard [a report that the wizard said something nasty].

(148) *What$_i$ did Dorothy hear [a report that the wizard said fl $_i$]?

Strong island effects can also be seen in subjects and adjunct clauses, as in the following examples.

(149) [A friend of the munchkins] told Dorothy to follow the yellow brick road.

(150) *Who$_i$ did [a friend of fl$_i$] tell Dorothy to follow the yellow brick road?

(151) Dorothy went to Emerald City [after she saw the Good Witch].

(152) *Who$_i$ did Dorothy go to Emerald City [after she saw fl$_i$]?

In (150), the extraction of *who* out of the subject renders the sentence ungrammatical, while in (152), the extraction of *who* out of the adjunct clause *after she saw* leads to ungrammaticality.

While these islands do block extraction in many cases, it is possible to leave a gap in an island if there is another grammatical instance of wh-movement in the proper configuration elsewhere in the sentence (so-called parasitic gaps). The examples in (153)-(156) show that gap sites can be present inside both adjunct and subject islands if there is another (grammatical) extraction elsewhere in the sentence (Kayne 1983).

(153) a man who [friends of *e*] admire *t*

(154) *a man who [friends of *t*] admire me

(155) a book which people buy *t* [without reading *e*]

(156) *a book that people understand linguistics [after reading *t*]

The examples in (157) and (158) show that parasitic gaps can only exist if they are embedded within a single island. If the gap is embedded within two islands, a grammatical movement elsewhere in the sentence does not change the ungrammaticality of the gap inside the islands (Kayne 1983, Richards 1997).

(157) *a man that I admire *t* [because [friends of *e*] become famous]

(158) *a book that people buy *t* [without understanding linguistics [after reading *e*]]

These movement facts have been analyzed a number of different ways (see Manzini (1992) for a clear summary). The key idea for a treatment of wh-movement is that the movement must be local all gaps must be locally bound by a c-commanding antecedent

(Chomsky 1973 and later). Long-distance movement is generally thought to be a series of local movements through specific landing sites (successive cyclic movement). Because movement must move through very specific sites, movement is blocked if another element is already in the movement site. The result of a filled movement site is a weak island (of the sort discussed in previous paragraphs), a constituent through which movement may not take place. One example of a configuration in which movement is blocked by an element in the movement site is an embedded clause that begins with a wh-element. Wh-elements are generally thought to occupy one of the locations through which movement must take place, [Spec, CP]. Thus, if an overt wh-element occupies that location, movement out of the clause is blocked.

Non-complement XPs (e.g. specifiers and adjuncts) define strong islands, the other important class of islands.

In what follows, I will adopt a theory reminiscent of Chomsky (1986), wherein certain categories block the search for an antecedent. In this work, CPs and DPs are assumed to be barriers to movement.

4.2 Parsing Moved Elements

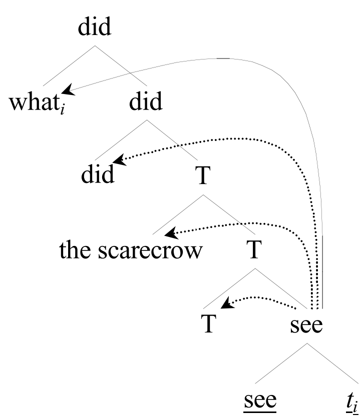
This section explains how intra-clausal wh-movement is accomplished in SPARSE, and then shows how long-distance (i.e. inter-clausal) movement is handled.

When a sentence-initial wh-element is encountered, it is not attached to anything, because there is no need to connect it to any other elements. However, when the next word is encountered, the [wh: +] feature on the wh-element allows it to be licensed in what has traditionally been called the [Spec, CP] position. When this attachment is made, the [wh: +] feature on the wh-element is checked, but because heads that license wh-elements do not license the other features found on NPs, the rest of the features of the wh-element remain unchecked. When a [wh: +] element is encountered, a wh-flag is set within the parser to indicate that there is an incomplete wh-chain in the sentence. Whenever a possible attachment site is encountered in the course of parsing and the wh-flag is set, a search is initiated for a c-commanding antecedent that can license the construction of a trace. For example, when the verb *see* is encountered in (159), its features allow for an NP object.

(159) What$_i$ did the scarecrow see *t* $_i$?

Because the wh-flag is set and there is a possible attachment site[34], the tree is searched for a c-commanding antecedent to see if a trace can be posited. In this case, the antecedent (*what*) contains features appropriate for the object position, so a trace is constructed. The process of positing the trace is shown in (160). When *see* is initially parsed, it triggers a search for an accusative argument. Because *what* has features appropriate for an accusative argument, a trace is posited as the direct object of *see*.

(160)



The search is initiated from the verb, and the trace that is posited as a result of the search requires an extra projection of the verb. The successful step of the search is marked with a solid line, the unsuccessful steps are marked with dotted lines. Newly-built material is underlined. Co-indexation of the wh-word and the trace indicates that the trace contains copies of all of the features from the antecedent wh-element. Any

[34] The definition of possible attachment site differs on the basis of the features of the wh-element. For example, the wh-element *why* contains a feature requiring that it select a verb on its left. Thus, whenever a verb is encountered, a trace of the wh-element will be attached to the right of the verb. NP arguments, such as *what*, on the other hand, require that case be assigned to them. Because of this, whenever a case-assigner is encountered, a trace will be posited.

feature that is copied from the wh-element to the trace is then checked on the wh-element. Thus, once a trace has been posited, all of the features on the wh-element are satisfied, and the features that were not satisfied on the wh-element are unsatisfied on the trace. This ensures that any features of the wh-element that were not satisfied by its initial attachment will be satisfied by one of its traces (e.g. if a wh-element does not receive case in its surface position, one of its trace must receive case).

Notice that because they are posited as soon as a possible licenser is encountered, many traces are posited that will need to be retracted later in the parse. For example, in (161), a trace is posited as the object of *see*, even though the direct object of *see* will end up being *him*.

(161) What did the scarecrow see __ him with?

The idea that traces are posited in advance of independent evidence for a gap has been given support in experimental work by Frazier, Clifton & Randall (1983), Crain & Fodor (1985), and Stowe (1986), and it has been dubbed the active filler strategy by Clifton & Frazier (1989).

Following work of Grimshaw (1986), I assume that maximal projections of C elements and of determiners (i.e. CPs and DPs) constitute barriers past which the antecedent search cannot continue. This is done to restrict movement out of the islands discussed in ⁄ 4.1. However, an absolute restriction on searching out of all CPs and DPs is too strong; it does not allow the long-distance movement out of a CP required for (162).

(162) What$_i$ did the scarecrow think [that Dorothy believed t_i]?

In order allow for long-distance movement within this system, it is necessary to construct intermediate traces in complement CPs and DPs when the wh-flag is set. This is accomplished in the following manner: whenever a new CP (or DP) is attached into a tree as complement of a verb or a preposition, a search is conducted for a c-commanding wh-phrase with unsatisfied features (which could therefore serve as antecedent to an intermediate trace in the Spec position of the new complementizer/determiner). If the search encounters a possible antecedent, a new trace is constructed in the lower [Spec, CP] position with the same features as the higher antecedent.[35] Because the CP (DP) itself triggers the search for an antecedent of movement, the barrier status of the CP (DP) triggering the search does not impede the search (i.e. the search does not move *through* the CP(DP) node).

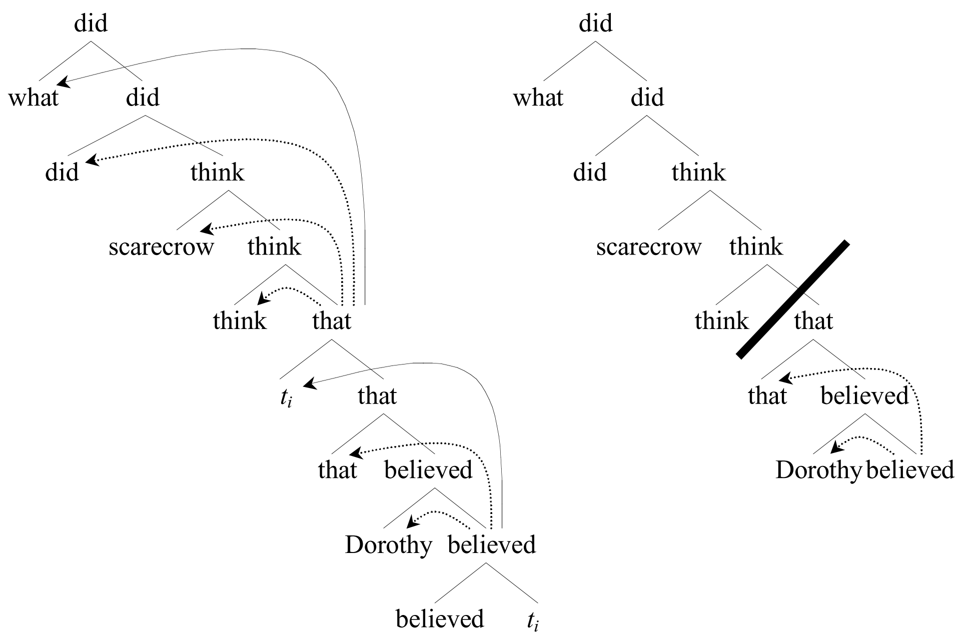
The tree on the left in (163) shows how successive cyclic movement is achieved. When *that* is attached to the existing tree, a search is begun for a c-commanding antecedent. When the search reaches the wh-element *what*, an intermediate trace is built in [Spec, *that*]. When *believed* is encountered, a search is initiated for an object antecedent (according to the normal gap-filling process discussed above), with the final trace built on the basis of the intermediate trace in [Spec, CP].

---

[35] Rather than searching the tree for a possible antecedent whenever a DP or CP is attached, the parser could pursue the option of stealing the trace that has already been posited (by active gap-filling) in the position where the DP/CP is attached (c.f. Pritchett 1991). However, there would still be instances when the search strategy would be needed, as in (vi) below, where a trace of *who* would never be inserted, because its features are not compatible the sentential complement requirement of the verb.

(vi) Who did she insist [*t* he would see *t*]?

(163)



The tree on the right in (163) shows that without the intermediate trace, the barrierhood of the complementizer *that* blocks the search necessary to find the antecedent of the direct object trace of *believed.*

The idea that cyclic movement of wh-elements passes through [Spec, CP] is well-established in the literature (Chomsky 1973, 1986, Kayne 1983). However, the idea that they also move through [Spec, DP] is less widely discussed. Definiteness islands are relevant to the question of whether or not wh-movement goes through [Spec, DP]. The basic fact is that definite NPs (as in (165), repeated from (145)) are more difficult to extract out of than indefinite NPs (e.g. (164)).

(164) Who did John find [a picture of *t*]?

(165) ??Who did John find [my picture of *t*]?

Melvold (1991) argues that this difference stems from the fact that there is an iota operator in [Spec, DP] of definite DPs, but not in indefinite DPs. The iota operator serves to bind the event position within the argument, thereby accounting for the referential effect of definiteness. If the presence of an element in [Spec, DP] can cause island effects, it is presumably for the same reason that an element in [Spec, CP] causes wh-islands because the movement is forced to pass through the specifier position by the fact that the DP (CP) is a barrier to extraction.[36]

In summary, this system effectively implements a successive cyclic analysis of long-distance wh-movement within an incremental system. The fact that this system works left to right makes it different from both the original conception of successive cyclic movement (Chomsky 1973) and its more recent incarnations (Chomsky 1986, Manzini 1992, Richards 1997). However, despite the differences in how the structure is arrived at, the basic structures produced by both this left-to-right system and the other bottom-up systems are very similar.

4.3 Limits on Wh-Movement

4.3.1 WH-Islands

The system for long-distance wh-movement outlined above provides an account of all the island effects on wh-movement discussed in ⁄4.1 . To begin, I will show why movement of adjuncts out of standard wh-islands is not possible in this system.

(166) *How$_i$ did Dorothy wonder which noises$_j$ the wizard could make t_j t_i?

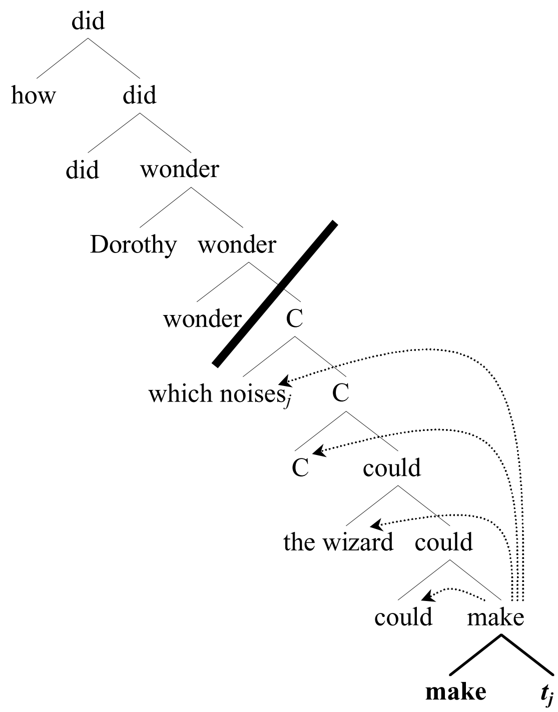
In (166), repeated from (139), at the point when the wh-word *which* in the embedded clause *which noises the wizard could make* is processed, it cannot be directly attached to the existing tree. Recall from ⁄ 2.5.2 every head has a distinguished feature that is used for the purposes of building predicted heads. In the case of wh-elements, the distinguished feature is [wh: +]. Thus, when there is no direct attachment site for *which*, the distinguished feature is used as the basis for predicting a licensing head. The only element that licenses a wh-feature is a C head. Thus, a C head is posited to license the

---

[36] I assume that indirect questions headed by *whether* are handled in a similar fashion. *Whether* is in the complementizer position, not in [Spec, CP]. However, I assume that a phonetically null element is in [Spec, CP] of these clauses, thereby rendering them weak islands to movement.

wh-element. On the basis of the selectional features of *wonder,* the C head is then attached as a complement of *wonder.* Ordinarily, the attachment of a C complement would trigger an attempt to insert a trace of cyclic movement. However, in this instance, [Spec, CP] is already filled by *which noises*, so no trace can be inserted. As can be seen in (167), when *make* is reached, a search for an antecedent takes place that reaches the wh-element *which noises*. Because *which noises* is a valid antecedent for an argument trace, such a trace is constructed in object position of *make*. However, when the time comes for the insertion of a trace for *how*, no such trace can be constructed. Because the wh-flag is set, the parser attempts to posit a trace for *how*. The search proceeds up to *which noises*, but because it does not have appropriate features for a VP-adjunct, no trace is constructed. Because the maximal projection of C is a search barrier, the search ends at that point, with no trace of *how* inserted. Thus, when the parse is finished, the wh-flag indicates that there is incomplete movement, and *how* still has not had its *[Cat: V, Left] feature checked.

(167)

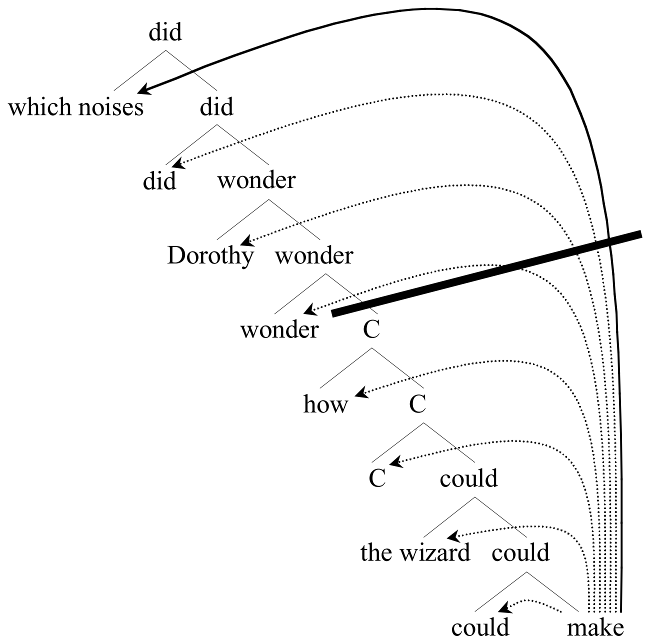


Recall from the discussion in ⁄ 4.1 that extraction of arguments out of wh-islands is marginally acceptable. For example, (168), repeated from (140), shows that extraction of the argument *which noises* is marginally acceptable.

(168) ??Which noises$_i$ did Dorothy wonder [how$_j$ the wizard could make fl$_i$ fl $_j$]?

This difference in acceptability is accounted for by allowing the search for antecedents of complements (but not non-complements) to continue past a barrier. When movement passes through a barrier, the features of the wh-element are satisfied, but the wh-flag that indicates incomplete movement is not changed to indicate completed movement.

(169)



The tree above shows the search process for the argument trace. Because the search for a complement trace antecedent is able to continue through a barrier (the question of which barriers the search can continue through will be addressed in the following section), a trace can successfully be built for the object of *make*. In both of these examples, the wh-islands are opaque to grammatical movement because [Spec, CP] cannot be filled by more than one item. If [Spec, CP] were to allow more than one element, it would be possible to posit an intermediate trace for the matrix wh-element, and the antecedent search would encounter a good antecedent without the need to pass through a barrier.

4.3.2 Adjunct Islands

Adjunct clauses are also islands for wh-movement, as can be seen by the fact that in (171), repeated from (152), and (173), neither the object *who* nor the adjunct *where* can be extracted.

(170) Dorothy went to Emerald City [after she saw the Good Witch].

(171) *Who$_i$ did Dorothy go to Emerald City [after she saw fl$_i$]?

(172) The wicked witch died [after the house landed on her in the munchkin village].

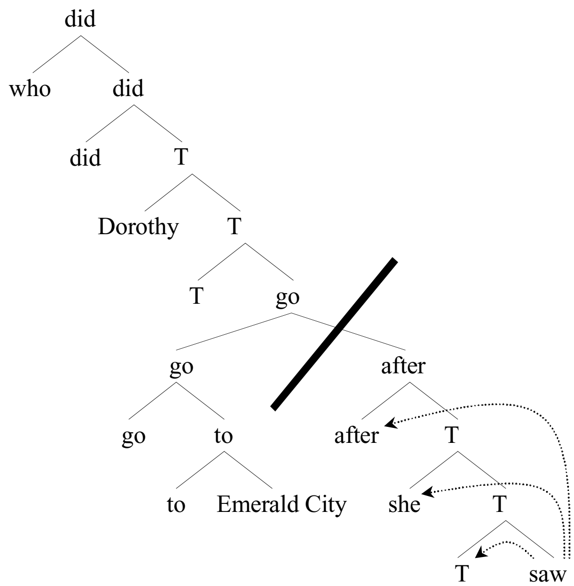
(173) *Where$_i$ did the wicked witch die [after the house landed on her fl$_i$]?

Recall that an intermediate trace is inserted in [Spec, CP] of complement CPs to serve as the necessary local antecedent of movement. However, in order to account for the fact that movement out of adjunct clauses is not allowed, traces of intermediate movement must not be posited in adjunct CPs. As noted above, the process of building intermediate traces is limited to complements. Thus, because the CP headed by *after* is an adjunct, it does not trigger the insertion of an intermediate trace.

In contrast to wh-islands, adjunct islands are strong islands not even arguments are allowed to extract out of them. Recall that it is marginally acceptable to extract arguments out of wh-islands; this is accounted for by allowing the search for an argument antecedent to proceed past barriers. To prevent the search from going past the adjunct barrier in (171), the antecedent search process will be modified slightly. In particular, if a maximal projection XP is a barrier and XP is a complement of a head Y, the antecedent search is allowed to proceed into YP. If XP is not a barrier, the search can always continue through to YP. This definition yields a search process that is able to ascend through complement barriers, but that is blocked by adjunct and specifier barriers.

As can be seen in the tree in (174), the search for an antecedent of an object trace for *saw* can only proceed up to *after*. Because the CP node is not a complement of *go*, the search does not continue up the tree and no antecedent is found.

(174)



The process of searching for an antecedent for an adjunct trace (e.g. a trace of *where* in (173)) proceeds along identical lines, with the search blocked because the search cannot proceed up through adjunct CPs.

4.3.3 Complex-NP Islands

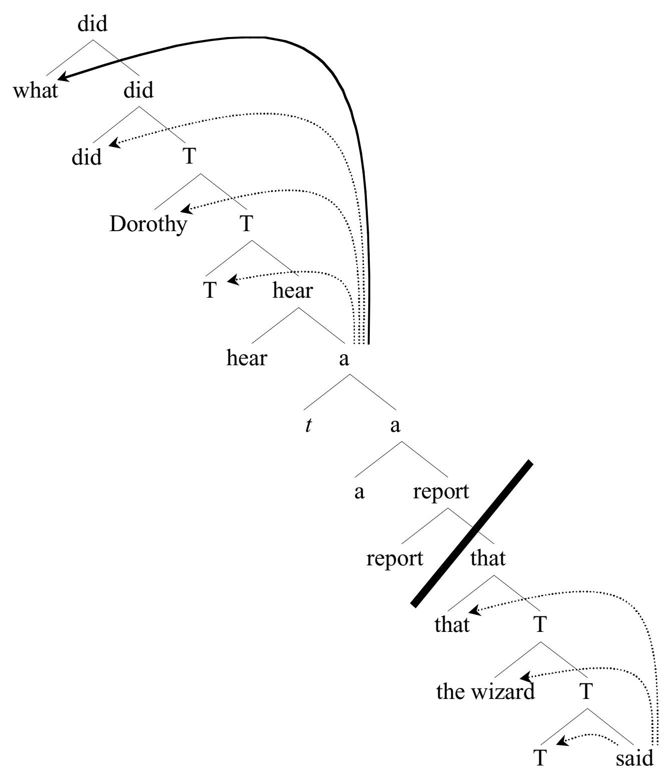
Complex NPs such as *the claim that the wizard said something nasty* in (175) (repeated from (147)) also constitute strong islands for extraction, as can be seen from the ungrammaticality of (176) (repeated from (148)).

(175) Dorothy heard [a report that the wizard said something nasty].

(176) *What$_i$ did Dorothy hear [a report that the wizard said fl$_i$]?

As can be seen below in (177), when the D^0 *a* is encountered, a search is carried out and a trace is posited in [Spec, *a*] However, when the complementizer *that* is encountered, no antecedent search is initiated. I rely on the analysis in Stowell (1981), in which the CP in a complex NP is an appositive (a type of adjunct), rather than a complement.[37] Because the CP is an adjunct, no search is initiated, and consequently no trace is placed in the specifier position of *that.* When *said* is processed, a search is initiated to determine whether or not a trace should be inserted as the object of *said*. The search proceeds up the tree to the complementizer *that*, where the search is blocked because of the adjunct attachment of the CP. Thus, no trace is inserted.

(177)



Because no object is found for *said* and because the features of *what* remain unchecked, the sentence is ungrammatical. Additionally, the wh-flag still specifies that there is an incomplete chain.

---

[37] This claim is based on the fact that the tensed clause complement in a complex NP can be equated with the NP itself, e.g. for the complex NP *Andrea s guess that Bill was lying,*

(vii) [Andrea s guess] was [that Bill was lying].

The idea is that a complement should not be able to enter into an identity relation with its predicate. Thus, the fact that the tense clause can enter into an identity relation with the NP is evidence that the clause is not an argument of the NP.

4.3.4 Subject Islands

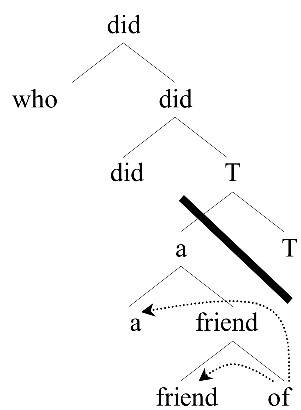
In (178) and (179) and, repeated from (149) and (150), we see that extraction from an otherwise well-formed subject is ungrammatical.

(178) [A friend of the munchkins] told Dorothy to follow the yellow brick road.

(179) *Who$_i$ did [a friend of t_i] tell Dorothy to follow the yellow brick road?

As can be seen in the tree in (180), the search for a possible trace antecedent can only proceed up as far as the DP node, because the DP barrier is not a complement. There is no trace in [Spec, *a*] because the determiner is not a complement and therefore does not trigger the search for an antecedent of movement.

(180)



Thus, the islandhood of subjects is accounted for by the fact that subjects are barriers but not complements, and they thus both obstruct the search process and fail to trigger the building of intermediate traces.

Sentential subjects (as in (182)) are subject to the exact same restrictions: the search is unable to proceed through the CP heading the subject, and no intermediate trace is present because subjects do not trigger an antecedent search.

(181) That the wizard had met the good witch surprised Dorothy.

(182) *Who did that the wizard had met surprise Dorothy?

4.3.5 Summary of Island Effects

As we have seen, movement is accounted for with a system in which traces of movement must be locally bound. An antecedent is locally accessible if there are no barriers between the trace and the antecedent. Maximal projections of complementizers and determiners are barriers. Long-distance movement is accounted for by the inclusion of intermediate traces in the Spec positions of the barriers. Islands are accounted for by disallowing intermediate traces. There are two different ways in which intermediate traces can be disallowed. Restrictions on the search process that triggers insertion of an intermediate trace (i.e. only search up through complements) account for subject, adjunct, and complex NP islands. The presence of elements in the location where the intermediate trace should be built account for islandhood in wh-islands, as well as definiteness islands. Extraction out of weak islands is allowed for complements because they are able to participate in a less restrictive search of the tree, while non-complements are more strictly limited. If this less restrictive search passes through a barrier, the wh-flag is not reset to indicate that movement has been completed.

4.3.6 Parasitic Gaps

To properly account for parasitic gap phenomena, the system for handling wh-movement must be modified slightly. The key to understanding parasitic gaps is to see that an extra gap can be present inside an island only if normal wh-movement elsewhere in the sentence creates a well-formed antecedent-trace relation. This generalization is translated into SPARSE in the following manner: the antecedent search is not stopped immediately upon reaching a non-complement XP (adjunct or specifier). Rather, it is able to continue on through one non-complement relation. If a trace is postulated from a search that passes through a non-complement relation, the wh-flag is changed to indicate that illicit movement has taken place. If a later instance of movement is successful (i.e.

passes through no barriers), the movement flag is changed to indicate that movement has been completed. At the end of the parser, if the wh-flag still indicates that movement is incomplete or illicit, the sentence is ungrammatical. This idea is similar to the Principle of Minimal Compliance in Richards (1997, 1998).

Consider what happens in sentences like (183) and (184).

(183) *a city that [citizens of *t*] like Emerald City

(184) a city that [citizens of *e*] like *t*

In both cases, when *citizens of* is parsed, the search for an antecedent to a trace of a prepositional object is forced to pass through the subject barrier (and the non-complement relation) at the DP headed by *citizens*. The wh-flag is set to indicate illicit movement when the trace is posited, because the search for the antecedent passed through a non-complement relation. When the end of the sentence is reached in (183), the flag has not been reset, so the sentence is ungrammatical[38]. When *like* is reached in (184), a search for an antecedent is initiated, and this results in the building of a trace as the object of *admire*. Because this is an instance of proper movement, the wh-flag is set to indicate completed movement. Thus, at the end of the sentence, the wh-flag does not indicate any incomplete or illicit movement, so there is no reason to judge the sentence ungrammatical. Notice that if the wh-flag indicates incomplete movement (as opposed to illicit movement) after the movement out of the subject in (183), movement out of a subject would have exactly the same characteristics as movement out a wh-island. However, movement of an argument out of a wh-island is noticeably more acceptable than movement out of a subject island. The difference between incomplete and improper movement values of the wh-flag accounts for this difference in acceptability.

An interesting question is when the parser stops actively positing gaps. If the parser continues to posit gaps as long as the wh-movement flag indicates some sort of

---

[38] Because of the active filler strategy, a trace is actually posited as the object of *like*, but it is removed when *Emerald City* is parsed.

movement (e.g. improper, proper, or incomplete), then the parasitic gap in (185), repeated from (155) above, should be filled actively. On the other hand, if gaps are only posited when the flag indicates incomplete movement, then the parasitic gap in (185) should not be filled automatically. Instead, it should only be filled when there is some confirmation that a gap is actually present.

(185) a book that people buy *t* [without reading *e*]

(186) *a book that people understand linguistics [after reading *t*]

I know of no existing experimental research to test whether or not parasitic gaps that occur after the proper gap are filled using the active-filler strategy.

As mentioned at the beginning of this section, the search for an antecedent is only able to pass through one non-complement relation. If a second non-complement relation is encountered, the search stops completely. By restricting the search in this way, the fact that parasitic gaps cannot be embedded within two or more islands can be accounted for. The examples in (187) and (188) (repeated from (157) and (158)) show that parasitic gaps can only exist if they are in a single island.

(187) *a man that I admire *t* [because [friends of *e*] become famous]

(188) *a book that people buy *t* [without understanding linguistics [after reading *e*]]

The fact that the search can only pass through one non-complement relation means that when *because* is encountered in (187), the search for an antecedent stops and there is no way to posit a gap as the object of the preposition *of*. Likewise, in (188) the search stops at *without*, meaning that no gap is posited after *reading*.

4.3.7 Multiple Wh-Fronting

The account of wh-islands presented above in 4.3.1 relies on the fact that only one wh-element is possible in [Spec, CP] in English (I will call this the wh-filter). In languages where more than one wh-element is allowed in [Spec, CP], wh-islands should not be respected. This section discusses evidence from Slavic and Balkan languages that shows that wh-islands are not respected precisely when multiple wh-elements are allowed in [Spec, CP].

Rudin (1988) shows that Bulgarian, Romanian, Serbo-Croatian, Polish, and Czech all allow fronting of multiple wh-elements.

(189) Koj kogo vi_da? (Bulgarian)
who whom sees
Who sees whom?

(190) Cine cu ce merge? (Romanian)
who with what goes
Who goes by what? (i.e. means of transportation)

(191) Ko koga vidi? (Serbo-Croatian)
who whom sees
Who sees whom?

(192) Kto co robi_ (Polish)
who what did
Who did what?

(193) Kdo koho videl? (Czech)
who whom saw
Who saw whom?

Despite this superficial similarity, Rudin presents a number of arguments that the [Spec, CP] position can be filled with multiple wh-elements in Bulgarian and Romanian, but not in Serbo-Croatian, Polish, or Czech. Rudin s arguments are summarized here the reader is referred to the original work for a full discussion. She argues that in Serbo-Croatian, Polish, and Czech, only the first wh-element is in [Spec, CP], while the others are inside the IP. Among her arguments for this analysis is the fact that only Romanian and Bulgarian allow multiple wh-elements to move out of complement clauses, as can be seen in (194).

(194) Koj k_de misli_ [_e e oti__l __ __] (Bulgarian)
who where think-2s that has gone
Who do you think that went where?

(195) *Ko _ta _elite [da vam kupi __ __]? (Serbo-Croatian)
who what want-2p to you buy
What do you want who to buy you?

(195) shows that Serbo-Croatian (like Polish and Czech) does not allow multiple extractions from within an embedded clause. I call this restriction the *wh-filter*. On the assumption that all movement from embedded clauses must pass through [Spec, CP], this provides an argument that only one element may be present in this position in these languages. Another argument comes from the fact that clitics in Bulgarian and Romanian appear after all the wh-elements, while in Serbo-Croatian, Polish, and Czech they appear after the first wh-element and before any others. Under the reasonable assumption that no elements are allowed to intervene between elements in the spec position(s) of a single head, this also provides an argument that the wh-elements in Serbo-Croatian, Polish, and Czech are not all in [Spec, CP]. Rudin further shows that Serbo-Croatian, Polish, and Czech allow the wh-elements to be interrupted by parentheticals, while this is ungrammatical in Bulgarian and Romanian. This fact is also consistent with an account in which only Bulgarian and Romanian allow the wh-elements to all reside in [Spec, CP].

Thus, it appears that the wh-filter is not operative in Bulgarian and Romanian, while it is operative in the other three languages. If Bulgarian and Romanian allow multiple elements in [Spec, CP], then the presence of one element in that position should not preclude another item from passing through that position. For this reason, wh-island effects are predicted to be absent from Bulgarian and Romanian. In accord with this generalization, Bulgarian and Romanian allow extraction from wh-islands, while Serbo-Croatian, Polish, and Czech do not.

(196) _oveka, kojito se _udi_ dali e do__l? (Bulgarian)
the person who wonder-2s whether has come
the person who you wonder whether has come

(197) * osoba, kojam sam ti rekao gde (on) _ivi (Serbo-Croatian)
individual who have-1s you told where he lives
the individual who you asked me where (he) lives

As can be seen in (196) and (197), Bulgarian (and also Romanian) allows relativization out of indirect questions, while Serbo-Croatian (and Polish and Czech) does not. Questioning out of relative clauses is also allowed in Bulgarian and Romanian. Questioning out of indirect questions is not possible in these two language, but Rudin attributes this to an unspecified non-syntactic restriction against questioning out of indirect questions. No extractions (relativization or questions) are possible out of either relative clauses or indirect questions in Serbo-Croatian, Polish, or Czech.

By deactivating the wh-filter for Bulgarian and Romanian, the SPARSE system can deal with these facts. Consider the example of extraction out of an indirect question presented in (196). When *dali* whether is attached to the existing structure, it triggers a search for an antecedent of wh-movement. Such an antecedent is found (the operator of the relative clause), and as a result, a trace can be posited in the Spec position. I assume that the trace is posited in a second Spec position, rather than being adjoined to the existing wh-element. The advantage of assuming a second Spec position, rather than adjunction to the existing Spec, is that with separate Specs, each can serve as a c-commanding antecedent to later traces[39].

As predicted by the model, removal of the wh-filter only affects wh-islands. Because the wh-filter only affects whether or not a wh-island can be voided via successive cyclic movement, its presence or absence has no effect on other islands. Richards (1997) presents evidence from Bulgarian showing that, in general, adjunct islands still serve as

---

[39] The addition of multiple specs requires one of two additional assumptions. Either the features of a head that license spec elements can be checked more than once (i.e. they can be checked by each spec), or space must be made in heads for the possible existence of two different sets of features for spec items. I do not believe that multiple specs requires any additional changes to structural assumptions.

effective barriers to wh-movement. This can be seen in the example below where *before we discuss* __ is an island that does not allow the wh-element *which question* to be moved.

(198) *Koj v_pros$_j$ iska Ivan da ka_e molitva [predi da obs_dim t_j]
which question wanted Ivan to say prayer before we-discuss
Which issue did Ivan want to say a prayer before we discuss?

4.3.7.1 Parasitic Movement in Bulgarian

Just as English parasitic gap constructions allow extraction out of adjunct islands, adjunct islands in Bulgarian are not completely resistant to extraction. Recall that a gap is allowed inside an English adjunct island if there is another instance of licit movement (i.e. in parasitic gap constructions). Bulgarian exhibits a similar phenomenon, in which extraction out of adjunct islands is allowed if there is another instance of valid wh-movement in the sentence[40].

(199) Koj profesor$_i$ koj v_pros$_j$ t_i iska da ka_e molitva [predi do obs_dim t_j]
which professor which question wanted to say prayer before we discuss
Which professor wanted to say a prayer before we discuss which issue?

In the example in (199), *which question* can be extracted out of the adjunct island *before we discuss* because of the presence of the valid extraction of the matrix subject. This can be handled in the same manner as English parasitic gaps. The wh-flag is set to indicate completed movement because of the proper movement of the matrix subject; there is therefore no reason to think that there is anything wrong with the sentence.[41]

One difference between Bulgarian and English is that Bulgarian parasitic movement involves two different overt wh-binders, whereas in English only one overt wh-binder is involved. However, the fact that there are two moved elements is not problematic in and

[40] There are certain instances when island violations cannot be rescued by valid movement elsewhere in the sentence. See Richards (1997) for a full discussion of the facts.

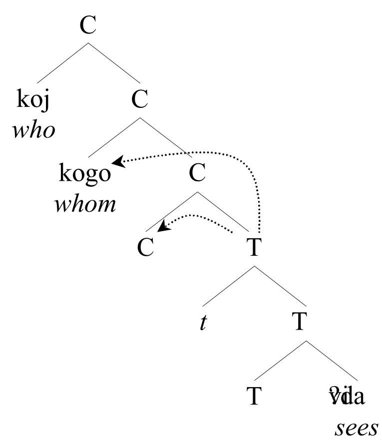
[41] Note that this assumes that improper movement does not change a wh-flag setting of proper movement to improper movement. It is assumed that the value cannot be changed once it has been set to proper movement.

of itself. When the second gap site is found (after *discuss* in (199)), the wh-element that is not already the antecedent of a trace (*which question*) is used as the antecedent.

While the number of moved elements is not problematic in Bulgarian, their ordering is. In general, the order of wh-elements in cases of multiple movement in Bulgarian is fixed. Except in cases of topicalization of one of the wh-elements, the wh-subject precedes the wh-object. This leads to a problem in determining which of the wh-elements should be the antecedent of which trace. Consider the simple example in (189), repeated here as (200). When the subject trace is built, the search for an antecedent for the trace should first encounter the accusative wh-element, as can be seen in (201).

(200) Koj kogo vi_da? (Bulgarian)
who whom sees
Who sees whom?

(201)



A trace of *whom* cannot be used for the subject in this case, because the case of *whom* is incompatible with a subject position. However, in the case-ambiguous example presented above in (199), there are no features to keep the second wh-element (the object *which question*) from being used as the antecedent of the subject. Accordingly, the method for searching for an antecedent must be modified slightly. In order to successfully account for the fact that the wh-subject is higher in the tree than the wh-object, the search will be modified so that it checks all the dependents of a projection, taking the highest one as the antecedent. Thus, in the tree in (201), the search would look at both *who* and *whom,* and would choose *who* as the first antecedent (for the subject) because it is in a higher projection of the C head. Likewise, in (199) *which professor* would be chosen over *which question* as the first antecedent because *which professor* is higher in the projection than *which question*. The fact that the highest antecedent is taken first in Bulgarian is the opposite of the English Path Containment Condition of Richards (1997, 1999). The reader is referred to Richards for an extensive discussion of the cross-linguistic differences in this.

4.4 Ambiguous Movement Structures

This theory of movement makes predictions for sentences in which it is not clear that movement has taken place (i.e. sentences that are temporarily ambiguous between a structure with movement and one without). The basic prediction is that the structures without movement will generally be preferred over structures with movement when it is not obvious that movement has taken place. Consider the examples in (202) and (203) (from Alphonce and Davis 1997).

(202) Ian is the man to watch.

(203) Ian is the man to watch Ardelia.

When *to watch* is parsed, all possible structures must include a PRO as the subject of *to watch*, but the structure might also include a null antecedent to movement from the object position (see the structures proposed by Browning 1987 in (204) and (205)). Because SPARSE generally only builds enough structure to combine the new word with the existing structure, the structure without the extra pro should be preferred.

(204) $[_{CP}$ pro$_i$ $[_{IP}$ PRO$_{arb}$ $[_{VP}$ $t_i]]]$

(205) $[_{IP}$ PRO $[_{VP}$ $]]$

If the structure in (205) is used, as predicted by SPARSE, then the sentence in (203) should be easier to process than (202). There should be added difficulty at the end of (202) because there is no object for the verb. The object position is not filled with a trace because the wh-movement flag has not been set, so the parser is not trying to fill any gaps. Furthermore, even if an antecedent search could be initiated, there should be no antecedent of movement to be found.

4.5 Summary

This chapter has shown how wh-movement is handled within the SPARSE system. An active-filler strategy is used in conjunction with an antecedent search to ensure that traces are inserted into the tree whenever possible. Barriers to the antecedent search, along with a wh-filter, ensure that islands are obeyed. Successive cyclic movement is invoked to account for the grammatical cases of long wh-movement. Parasitic gaps are accounted for by allowing traces to be posited if there is one intervening barrier, as long as another successful movement validates the movement. In order to account for complement/non-complement asymmetries in weak islands, complements are allowed to use a slightly more permissive search mechanism. The account is shown to be compatible with findings from languages which do not have the wh-filter. Finally, the system predicts that in cases where it is temporarily unclear whether there is movement, the preference should be to assume that there is no movement.

4.5.1 Relation to Incrementality and Grammar

The theory of wh-movement discussed in this chapter can be seen as an essentially autonomous module of an incremental parser. The main ideas of this system are not tightly linked to the details of incremental parsing presented in the previous chapters. Instead, the ideas proposed here could be relevant to a wide variety of bottom-up incremental parsers, and could probably also be used to advantage in a top-down parser as well, though the details would likely require changes.

This system has not been included in the parsing algorithm discussed so far. However, I believe that the system has been explained clearly enough that it could be implemented fairly straightforwardly, as a simple add-on to the main SPARSE parsing algorithm.

One question that frequently arises in discussions of this work is its relationship to the grammar. Is this account supposed to take the place of a grammatical account? Is it supposed to show that wh-movement constraints really boil down to resource limitations or that they are related inextricably to some incremental parsing phenomenon? Pritchett (1991) argues that constraints on wh-movement are not grammatical in nature, but are instead the result of incremental processing. The constraints on reanalysis that are in effect throughout the parsing process are argued to be responsible for the fact that extraction out of islands is impossible.

In contrast to Pritchett, I do not view an incremental treatment of wh-movement as obviating the need for a grammatical analysis of movement constraints. Instead, I see this work as an example of how movement (and constraints on movement) can be translated from a normal bottom-up syntactic analysis to an incremental left-to-right syntactic analysis. As such, constraints on movement are still very much grammatical in nature.

This account of wh-movement also represents an attempt to show how some ungrammatical constructions might be parsed in an incremental system. The fact that traces are sometimes posited even when they are not grammatically licensed (i.e. they don t complete a chain or are example of illicit movement) provides the mechanism necessary for processing and interpreting ungrammatical instances of movement. I see this as the beginning of a larger attempt to show how ungrammatical constructions can be parsed and understood. The key to parsing these constructions is to slightly relax the constraints that typically guide parsing. I suspect that this general approach can also be applied to other types of ungrammaticality, conceivably even instances of temporary ungrammaticality that are eventually made grammatical (like the parasitic gap cases discussed above). In this sense, this work is in the spirit of the Diagnosis Model proposed by Fodor and Inoue (1994, 1998), in which they propose that ungrammatical attachments are sometimes made when no other options are available.

Chapter 5

EXPERIMENTAL RESULTS

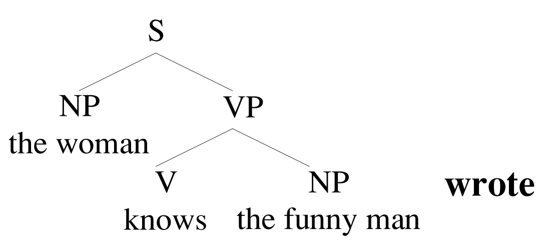
5.1 Introduction

One important question related to the parsing algorithm is that of how the grammatical search space is navigated. Specifically, are all local attachments attempted before any less local attachments are attempted, or are some non-local attachments considered before all of the local attachments are exhausted? Consider the following example of the well-known NP-S ambiguity in which an NP (*the funny man*) can be either a direct object NP or the subject of an embedded clause:

(206) The woman knows the funny man wrote

Under standard assumptions, the structure of this sentence before *wrote* is processed is that of a simple transitive sentence with *the funny man* serving as the direct object of *knows*.

(207)

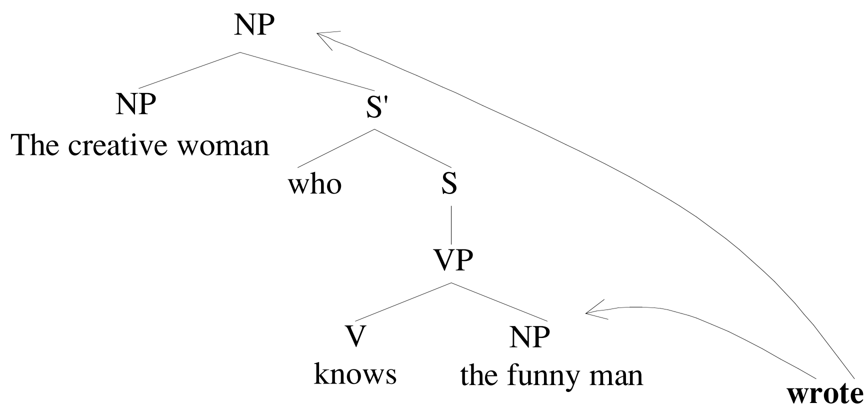


When the verb *wrote* is encountered, the NP *the funny man* must be made the subject of an embedded clause headed by *wrote*, with the embedded clause serving as the complement of *knows*. Because there is a unique solution, it is obvious what the final structure must be. However, it is less obvious how this structure is arrived at. Since the NP *the funny man* is parsed as the direct object of *wrote*, there might be a preference to respect that existing commitment and search for other possible attachments of *wrote* before changing the role of the NP *the funny man* in the sentence. If there is a preference to respect existing grammatical commitments, other options should be searched for before existing commitments are broken. In this example, this means that the part of the tree above *knows* should be searched for possible attachments of the new verb before *the funny man* is changed to be the subject of a sentential complement. This strategy of respecting existing commitments until there are no other options has been called Reanalysis as a Last Resort (Fodor and Frazier (1980), and see Frazier (1990) and Frazier and Clifton (1998) for a discussion of Minimal Revision , a similar constraint).

On the other hand, if locality is a decisive factor in parsing, the most local attachments should be attempted before any less local attachments are attempted. In this case, this means that the part of the tree above *knows* should not be searched until all possible local options have been exhausted. Thus, *the funny man* should be analyzed as the subject of *wrote* before any other parses involving material above *the funny man* are pursued. This hypothesis will be called Locality First . The idea that the non-local attachments are not even considered in sentences like (206) has been lent credence by the fact that the structural change required for this attachment has been shown to be very easy (Sturt et al. 1999b). To account for the easy processing of this ambiguity, some theories (Marcus et al. 1983, Weinberg 1993, Gorrell 1995) have even claimed that this change requires no retraction whatsoever of existing structural commitments, i.e. it doesn t require any reanalysis. The question of whether reanalysis is a last resort operation can be easily overlooked in studies of ambiguity resolution, because sentences like (207) are unambiguous. However, this question must be addressed in implementations of incremental parsers, otherwise there is no way to determine whether or not the sentence in (207) is ambiguous.

To test these hypotheses, two self-paced reading experiments were performed. The experiments involved embedding the NP-S ambiguity shown in (206) above in a context where there are two possible subjects for the verb *wrote*. To provide these two different attachment sites, the standard NP-S ambiguity (which provides one possible attachment site) was embedded in a subject relative clause on a matrix subject. The matrix subject provides the other possible attachment site. An example of this temporary ambiguity can be seen in (208).

(208)



In the tree above, the two possible subjects for the incoming verb are illustrated. The verb that can be attached in two different places (*wrote*) will be called the ambiguous verb , and for reasons that will be explained later, the NP-S ambiguous verb will be called the embedding verb . The matrix subject NP (*the creative woman*) will be called the high NP , while the NP object/embedded subject (*the funny man*) will be called the low NP . Likewise, the structure in which the high NP (*the creative woman*) is the subject of the ambiguous verb (*wrote*) will be called the high condition/attachment , while the structure in which the low NP (*the funny man*) is the subject of the ambiguous verb will be called the low condition/attachment (see trees in (211) below).

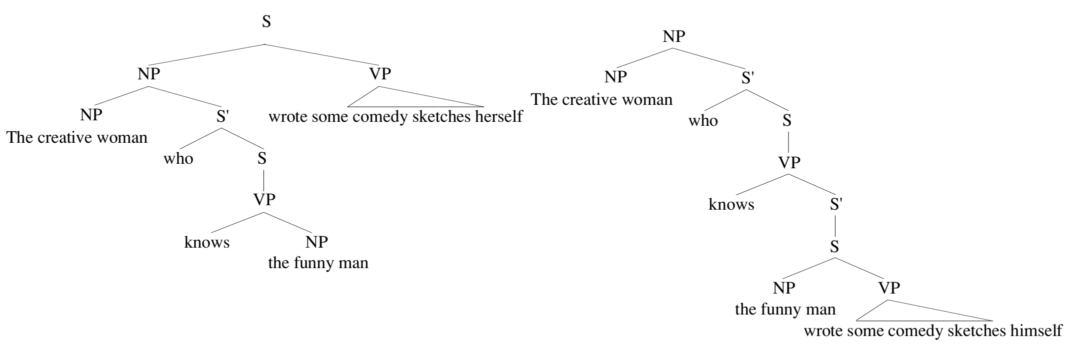
To disambiguate the sentences, a gender-marked emphatic reflexive was included four words after the ambiguous verb. Emphatic reflexives, like other reflexives, must be locally bound by an agreeing antecedent. The sentences in (209) and (210) demonstrate that antecedents for the emphatic reflexives must be within the same clause as the reflexive.

(209) Dorothy$_i$ knew [that the tinman$_j$ chopped down the trees himself$_j$].

(210) *Dorothy$_i$ knew [that the tinman$_j$ chopped down the trees herself$_i$].

The logic of the experiment was as follows: difficulty at the disambiguating reflexive relative to the unambiguous conditions should be indicative of an earlier parsing error. Thus, if a low attachment strategy is followed (as in the tree on the right in (211)), *himself* should cause no difficulty. In contrast, *herself* should cause difficulty, since it cannot be properly bound by a clause-mate NP. On the other hand, difficulty at *himself* indicates that the high attachment has been made, as in the tree on the left, since *himself* can t be locally bound by the intended masculine antecedent (*the funny man*).

(211)



Each experimental item contained four conditions, as shown below:

(212) *Low Ambiguous*
The creative woman who knows the funny man wrote some comedy sketches himself about the amusing escapades thinks he should publish them.

(213) *Low Unambiguous*
The creative woman who knows that the funny man wrote some comedy sketches himself about the amusing escapades thinks he should publish them.

(214) *High Ambiguous*
The creative woman who knows the funny man wrote some comedy sketches herself about the amusing escapades she had seen.

(215) *High Unambiguous*
The creative woman who knows him wrote some comedy sketches herself about the amusing escapades she had seen.

The first and second (high and low) NPs of the sentences were all human NPs, either male or female. In each sentence, the gender of the two NPs differed, with either a MASC-FEM order or FEM-MASC order. The orders were counterbalanced across items. Within each experimental item, there was a low ambiguous condition (e.g.(212)) in which the reflexive agreed in gender with the embedded subject (the low NP), and a high ambiguous condition (e.g.(214)), in which the reflexive agreed with the matrix subject (the high NP). In addition to the ambiguous conditions, unambiguous controls were also included. The controls in the low attachment condition of Experiment 1 (e.g. (213)) were disambiguated by including the complementizer *that* immediately prior to the low NP. This had the effect of forcing the following NP to be the subject of a complement clause. The high controls (as in (215)) were disambiguated by using an accusative pronoun in place of the low NP. The accusative case-marking on the pronoun forced it to be attached as the direct object, and prevented its later reanalysis as the subject of an embedded clause.

If the Locality First search strategy is used and low attachments are made in ambiguous conditions, there should be an increase in reading times for the high ambiguous conditions relative to the high unambiguous conditions. This increase would be seen at the disambiguating reflexive, because the reflexive would be incompatible with the low attachment posited by the parser. This incompatibility would then trigger a revision of the initial low attachment, with this revision process reflected in a slowdown at the disambiguation. Under Locality First, there should be no comparable slowdown in the low ambiguous conditions, since the parser should already be pursuing the local analysis (low attachment) in ambiguous conditions. Thus, there should be no surprise/slowdown upon seeing a reflexive consistent with the low attachment.

Contrasting with the predictions of Locality First are the predictions of Reanalysis as a Last Resort (RALR). RALR predicts that there will be no slowdown in the ambiguous high conditions relative to the unambiguous high conditions, since the high attachment of the verb should be pursued in all ambiguous cases. For the low conditions, RALR predicts that the disambiguating reflexive will be read more slowly in the ambiguous conditions than in the unambiguous conditions. Under RALR, a slowdown in the low ambiguous condition is accounted for by the need to revise the initial high attachment to the lower (embedded clause) attachment site.

5.2 Experiment 1

The experiments reported here were performed in collaboration with Colin Phillips, and were presented at the CUNY Human Sentence Processing Conference in March 1999 (Schneider and Phillips 1999). Experiments very similar to Experiment 1, performed independently of these, were also presented at CUNY 99 by Patrick Sturt and colleagues (Sturt, Pickering and Crocker 1999a). The bare NP-S ambiguity has been studied by Frazier and Rayner (1982), Ferreira & Clifton (86), and Trueswell, Tanenhaus and Kello (1993) among others. I am aware of no other studies besides that of Sturt, et al. (1999a) that have tested RALR experimentally.

5.2.1 Materials

Experiment 1 included 48 sets of experimental items, each as described in the preceding section. It has been shown (Trueswell et al. 1993) that the relative frequencies of different complements in corpora have effects on processing speed/difficulty. Specifically, Trueswell et al. (1993) showed that in a sentence like (216) (repeated from (206) above), the verb (or the words immediately following the verb) is read more slowly when it follows an embedding verb that most frequently takes NP complements than when it follows a verb that most frequently takes sentential complements.

(216) the woman knows the funny man wrote

Because these effects of verb-complement bias could potentially obscure any otherwise relevant effects, the verbs used in the experiment were chosen to belong to two homogeneous groups. 24 of the items contained strongly NP-biased verbs (83%-100% NP completions), while 24 contained weakly NP-biased verbs (52%-78% NP completions). S-biased verbs were not included in this experiment for several reasons. First, it is possible that NPs following strongly S-biased verbs are initially attached as subjects of embedded clauses rather than as direct objects (as suggested by Trueswell et al. 1993). A low attachment of the ambiguous verb in such a sentence would therefore not require reanalysis. This would thereby eliminate the trade-off between locality and reanalysis that the experiment is built around. Previous research has also shown that there are strong locality effects in parsing (Kimball 1973, Wanner 1980, Gibson et al. 1996, Phillips and Gibson 1997a,b, Gibson 1998 and others).

High attachment controls were included in the study because of limits on the interpretability of low attachment findings. In particular, low attachments could be the result of RALR simply being a weaker constraint than locality. Thus, low attachments would be insufficient to rule out the possibility that RALR is a real parsing constraint/strategy that is simply over-ridden by a more powerful locality effect. On the other hand, high attachments would provide strong evidence that RALR is guiding the parser, because such a finding would indicate that RALR is able to override locality effects.

If the ambiguous NP is initially parsed as subject of an embedded clause, the question of how to interpret the results becomes much less clear, because the local (low) attachment does not require any reanalysis. Because S-biased embedding verbs might cause the ambiguous NP to be initially parsed as a sentential subject, these verbs were not included in Experiment 1.

The inclusion of two classes of NP-biased verbs resulted in a 2 x 2 x 2 design, with the following factors: ambiguity, attachment type (high/low), and verb class. The items consisted of subject relative clauses attached to matrix subjects, followed by a temporarily ambiguous verb. The high NP and the low NP were animate NPs of opposite gender. Four words downstream from the ambiguous verb was a gender-marked emphatic reflexive that disambiguated towards either a high or low attachment. Immediately following the reflexive was a four-word PP. Additional material followed the PP to make the sentences more natural and to ensure that the measurements from the PP were not influenced by the slowdown generally found at the end of the sentence. In the low conditions, this extra material included the matrix verb necessary to make the sentences grammatical. In the low conditions, the four-word PP also served as a buffer so that any effects of the reflexive could be measured before the verb in region 13 that must be attached high (i.e. in a non-local position). An example set of stimuli can be found above in (212)-(215).

An example of the region encoding (for the sentence in (213)) follows:

(217) The creative woman who knows / that / the funny man / wrote / some / comedy /
1 / 2 / 3 / 4 / 5 / 6 /
sketches / **himself** / about / the / amusing / escapades / thinks / he should publish
7 / 8 / 9 / 10 / 11 / 12 / 13 / 14
them.
14

The verbs used in the experiment all allowed both NP-complements and sentential complements. An additional constraint was placed on the verbs they were all required to allow animate NPs as direct objects. Table 1 shows the verbs used in the experiment, along with the data used to determine the bias classification. The data come from completion studies performed by Susan Garnsey (p.c.) and are similar to the more limited results in Trueswell, Tanenhaus and Kello (1993). The first two columns of data show the number of sentences in which subjects completed the sentence fragment using either an NP direct object or a *that*-less sentential complement. These two types of completions are similar to the ambiguous conditions in the experiment in that after the first NP following the verb, they are still compatible with either an NP-complement analysis or an S-complement analysis. The final column shows the percentage of these ambiguous completions that were completed with sentential complements. In addition to the *that*-less sentential complements, there were also many S-complement completions that included the complementizer *that*. These completions were excluded from the calculations used to determine verb class, since the goal of this experiment was to find out what subjects do in the ambiguous conditions. Including the

S-complements with *that* would have resulted in the same groupings, though with higher percentages (generally 15%-20% higher) of S-complements.

Table 1: Sentence Completion Data for Verbs in Experiment 1
From Susan Garnsey (p.c.)

	NP completions	***that*-less S-complements**	**% S-complements**
Weak NP-bias			
know	31	29	48%
doubt	42	20	32%
mention	41	17	29%
fear	32	11	26%
notice	58	16	22%
Strong NP-bias			
discover	69	9	12%
acknowledge	69	5	7%
hear	74	3	4%
warn	74	2	3%
appreciate	95	0	0%
understand	72	0	0%

An objection might be raised that the verbs *appreciate* and *understand* are not really ambiguous, since there were no *that*-less S-complements. However, both of these verbs did appear with S-complements containing *that* and they were judged by a number of informants to be grammatical with *that*-less S-complements, as shown in the examples below.

(218) I understand he can do the job

(219) They appreciated he was able to unlock the door for them

101 filler sentences were also included in the experiment. 36 of the fillers were experimental sentences from an unrelated experiment that was being run concurrently; the other 65 were distracters. The fillers from the unrelated experiment were of similar length to those in this experiment. The distracters were of varied length, ranging from shorter than the experimental sentences to longer than the experimental sentences. The

order of presentation of the sentences was randomized across subjects. Subjects saw only one condition of each stimulus item.

5.2.2 Method

The materials were presented to subjects on a computer screen using the word-by-word self-paced moving-window paradigm (Just et al. 1982). The size of the window was one word. The portions of the sentences before and including the disambiguation all appeared on the same line. Some of the remainder of the stimuli appeared on the next line. The trials were presented in a single block, with a yes/no comprehension question following each item. The majority of the comprehension questions were about the ambiguity, but the questions were varied somewhat to keep the subjects from adopting strategies specific to the experimental sentences. A version of the Daneman & Carpenter reading span task (Daneman and Carpenter 1980) was presented to each subject prior to the experiment. In this task, the subject reads a sentence aloud to the experimenter and then answers a yes/no question about the just-read sentence. After doing this for a group of sentences, the subject is asked to recall the last word of each of the sentences that were read out loud. The reading span score is equal to the size of the sentence groups for which the subject is able to successfully recall all of the final words in 4 out of 5 trials plus .2 points per successful trial in the next size group. For example, if a subject successfully completes four of the five trials correct in groups of two sentences, and also completes two trials in the three sentence condition, the subject s score would be 2 (size of largest successful group) + 2 (number of trials of size 3) * .2 = 2.4.

A regression equation was run on reading times to factor out the effects of word length. A regression equation was constructed for each subject to predict reading time for words of different lengths. All items (filler and experimental) were used in constructing the regression equation. The residual reading time is determined by subtracting the reading time predicted by the subject s regression equation from the actual measured reading time. This transformation removes extraneous variance by subtracting out a baseline for each subject, and by controlling for noise due to word-length effects (Ferreira and Clifton 1986, Trueswell and Tanenhaus 1991). All times reported here are average adjusted residual reading times of the trials for which the subject answered the comprehension question correctly. Reading times that were more than 4 standard deviations from the mean were trimmed to 4 s.d. (1108 ms.). This effected 411 of 43962 total words (0.93%). Trials for which subjects answered the comprehension question incorrectly were excluded from the analyses below.

5.2.3 Subjects

63 subjects from the University of Delaware community participated in the study. All were paid for their participation. Subjects were eliminated from the analysis if they scored less than 90% on the comprehension questions on the distracters, or if they scored less than 75% on the experimental questions. 15 subjects were eliminated from the analysis due to low comprehension scores, leaving 48 subjects

5.2.4 Results

In the discussion of the results that follows, the subjects will be split into two different groups based on their reading-span test scores. The reading-span test was run as part of the other experiment being run concurrently with the one reported here, and was not expected to have any bearing on the results from this study. However, because there were effects of reading span for the other experiment, a split analysis was performed on this experiment which yielded significant differences between the two groups. For this reason, the discussion will be divided according to reading-span group.

The reading times at the disambiguating region will be discussed first to establish what the ultimate attachments made by the subjects are. The earlier regions will then be discussed to see if they can illuminate the details of the parsing process. The analyses performed were repeated measures ANOVAs with region, ambiguity, verb class, and attachment site as within subject factors.

Due to an unusually flat response, the trials with *fear* were excluded from the analysis. One condition of one item (5b) was removed from the analysis due to a typographical error that rendered the sentence ungrammatical. As a result, the final analysis includes 19 items in the weakly NP-biased condition and 24 in the strongly NP-biased condition.

5.2.4.1 High Span Subjects

High span readers were defined in this experiment as those having a reading-span test score of 2.6 or greater (n=22).

5.2.4.1.1 Reading Times

The two different verb classes are combined in the graphs of the high span subjects for ease of exposition, since there were no main effects of verb class nor any interactions involving verb class for these subjects. Figure 5 shows the average residual reading times of the high memory-span subjects in the conditions requiring low attachment.

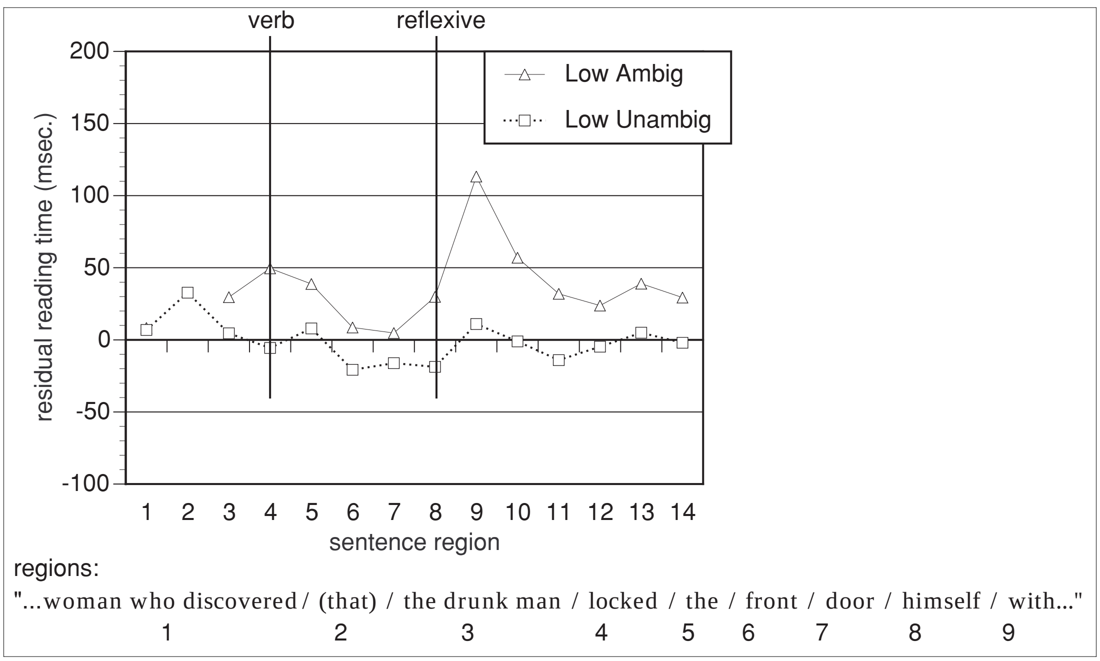
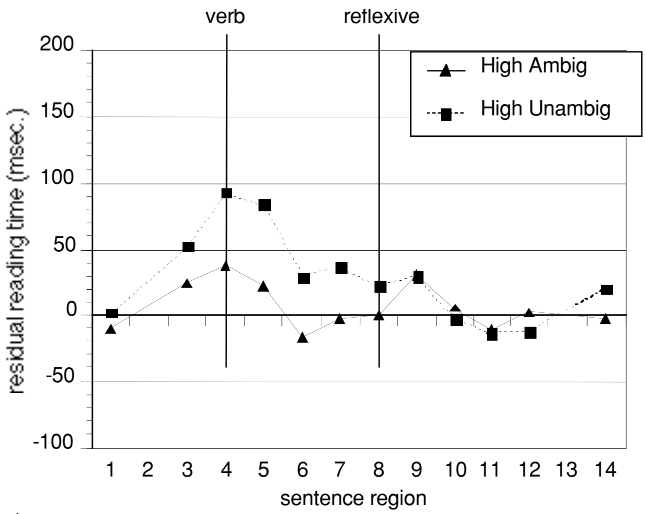


Figure 5: Expt. 1, High Span Subjects, Low Conditions, n=22

At the word immediately following the reflexive (region 9), there was a significant effect of ambiguity ($F_1(1,21)=22.4$, $p<.001$, $F_2(1,41)=22.5$, $p<.001$). The fact that this effect occurred one word after the disambiguation is typical of word-by-word reading experiments the effects of a word are frequently found one or two words downstream of the word causing the effect. There was no main effect of verb class, and no verb class x ambiguity interaction (all Fs<1). The main effect of ambiguity was also significant at region 8 ($F_1(1,21)=5.06$, $p<.05$; $F_2(1,41)=6.86$, $p<.01$), region 10 ($F_1(1,21)=8.89$, $p<.01$, $F_2(1,41)=9.59$, $p<.01$), and region 11 ($F_1(1,21)=7.56$, $p<.01$, $F_2(1,41)=6.65$, $p<.05$), but was not significant at region 12 ($F_1(1,21)=2.58$, $p<.11$; $F_2(1,41)=2.24$, $p<.14$). The main effects at region 8-11 are most likely all due to the reflexive and the revision processes associated with it, with the revision process still having some effect as late as region 11.

Given that there was a significant effect of ambiguity throughout the entire sentence ($F_1(1,47)=54.2$, $p<.0001$; $F_2(1,42)=59.1$, $p<.0001$), the demonstration of an ambiguity effect at the disambiguation (see previous paragraph) is not the strongest possible evidence that the two conditions involve different computations at the disambiguation. It could be the case that the general cost in the ambiguous condition is responsible for this difference. However, the fact that there was an interaction between region and ambiguity precisely at the point of disambiguation (regions 7-9, $F_1(2,21)=4.10$, $p<.05$; $F_2(2,41)=4.17$, $p<.05$) provides strong evidence that the difference between the two conditions is more than just the ambiguity cost seen over the entire sentence. This interaction did not reach significance in region 8-10 or 9-11 (all p>.1). I interpret this interaction to mean that the reflexive causes a differential load on the parser at the point of disambiguation (i.e. the reflexive is more difficult in the ambiguous conditions than the unambiguous conditions). Thus, this interaction with ambiguity provides evidence that the subjects sometimes make the high attachments and are then forced to revise them.



regions:
"...woman who discovered / the drunk man / locked / the / front / door / herself / with / the... "
1 3 4 5 6 7 8 9 10

Figure 6: Expt. 1, High Span Subjects, High Conditions, n=22

In the high conditions shown in Figure 6, there were no significant main effects or interactions at region 9 (all Fs<1). There were no interactions between ambiguity and region in any of the regions near the ambiguity (i.e. no interactions at 7-8, 7-9, 8-9, 8-10, 9-10, 9-11 (all Fs<1.3)).

The fact that there was a main effect of ambiguity at the disambiguating regions in the low conditions, but not in the high conditions suggests that in the ambiguous conditions, the high attachment is always pursued. As mentioned earlier, the increase in reading time in the low ambiguous conditions is a consequence of the need to repair the initially-pursued high attachment. If low attachments had also been pursued (at least sometimes), there should have been an effect of ambiguity in the high conditions, since the low attachments would have needed to be revised.

Given that the ambiguous conditions appear to initially receive a high attachment (based on the evidence at the disambiguation), the discussion will now turn to the earlier ambiguous regions. In the low conditions, the NP at region 3 (immediately prior to the ambiguous verb) was read faster in the unambiguous condition than in the ambiguous condition ($F_1(1,21)=3.76$, $p<.06$, $F_2(1,41)=5.60$, $p<.05$). This effect is apparently due to the fact that an NP is unambiguously predicted in the unambiguous condition. Recall that in the unambiguous condition there is a complementizer *that* immediately prior to the NP which provides a very strong cue that an NP follows immediately. In contrast, in the ambiguous condition, the embedding verb (that immediately precedes the NP in region 3) could be followed by either an NP or a complementizer.

At the verb itself (region 4) in the low conditions, there was a significant main effect of ambiguity ($F_1(1,21)=7.74$, $p<.01$, $F_2(1,41)=8.69$, $p<.01$), with the ambiguous conditions being read more slowly than the unambiguous conditions. There were no other main effects or interactions (all Fs<1). I would suggest that the verb in the low ambiguous condition was read more slowly than the low unambiguous condition because of the cost of searching the tree for the higher attachment site. In the low unambiguous condition, the complementizer *that* has already signaled that the lower, more local attachment of the verb must be made. Thus, no search need be done in the unambiguous conditions. In the ambiguous conditions, I have already shown that the verb is attached high (see the evidence above from the disambiguation), and this attachment must be made precisely when the verb is encountered. The search for the higher attachment site can easily account for the increased reading time in the low ambiguous conditions.

In the high conditions, there was no significant main effect of ambiguity at the NP in region 3 ($F_1(1,21)=2.48$, $p<.12$; $F_2(1,41)=2.57$, $p<.11$), though there was a numerical tendency for the unambiguous conditions to be read more slowly. Given that the unambiguous condition contained a three-letter pronoun, while the ambiguous conditions contained longer full NPs, this tendency could be due to word-length effects that were not completely factored out by the regression equation, or it could be due to effects that might have been created by the regression equation. One other possible explanation is that the numerical effect is due to the need to search for an antecedent for the pronoun. There was neither a main effect of verb class nor a verb class x ambiguity interaction (all Fs<1). This is consistent with the fact that the NP is being analyzed as an object NP in both conditions.

In the high conditions, there is a main effect of ambiguity at the ambiguous verb (region 4) ($F_1(1,21)=5.01$, $p<.05$, $F_2(1,41)=4.32$, $p<.05$), although in contrast to the low conditions, the *un*ambiguous trials were read more slowly than the ambiguous trials in the high conditions. There was no main effect of verb class, nor was there a verb class x ambiguity interaction (all $Fs<1$). The slowdown in the unambiguous conditions in region 4 is consistent with a delayed search for the antecedent of the pronoun (which provides the early disambiguation). Recall that in region 3, there was also an increase in reading times for the unambiguous high conditions relative to the ambiguous high conditions, although the difference was not significant. The fact that the slowdown was not significant until the word after the NP is likely due to the phenomenon commonly found in self-paced reading studies of downstream effects , in which effects of difficulty at one word are seen at later word. In addition to requiring an antecedent search, the lack of an antecedent for the pronoun renders the sentences somewhat unnatural, which might account for some of the difficulty in the high unambiguous conditions (this was corrected in Experiment 2). Given that there is no evidence from the reading times at the point of disambiguation that the attachments initially pursued in the ambiguous and unambiguous conditions are any different, I do not think that this difference between ambiguous and unambiguous conditions can be attributed to the search for an attachment site. In fact, the cost for searching for the higher attachment should be present in both the high ambiguous and high unambiguous conditions, since there is evidence that the high attachment is made in both conditions.

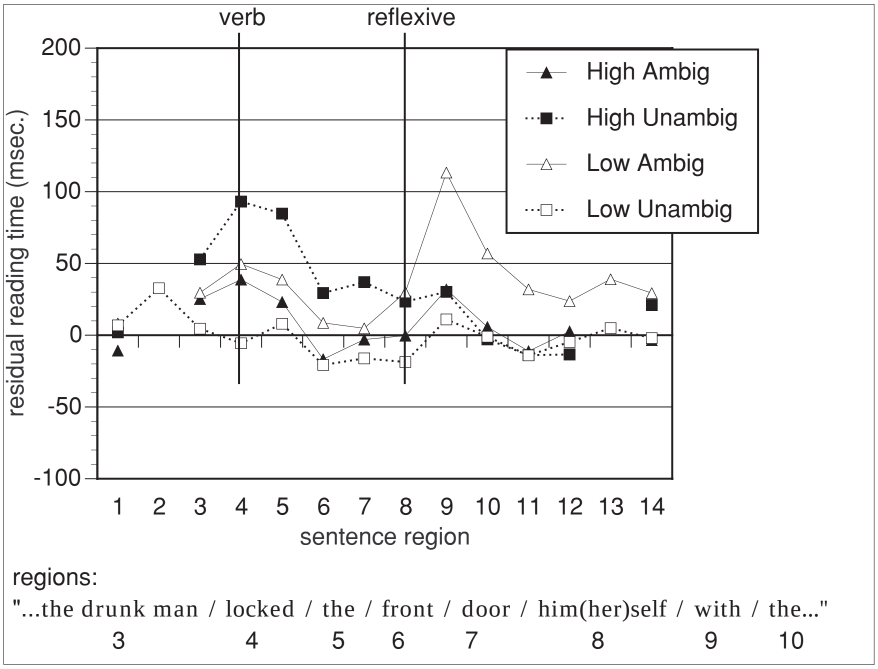


Figure 7: Expt. 1, High Span Subjects, Verb Classes Combined, n=22

In the high and low ambiguous conditions (the solid lines in Figure 7), the words up to the reflexive are exactly the same. As can be seen in the graph showing all the conditions, there seems to be a difference in the reading times between the high and low ambiguous conditions, despite the fact that the words in these regions are identical. When regions 1-7 (the entire identical region) were combined, the low ambiguous conditions were read significantly more slowly than the high ambiguous conditions (($F_1(1,21)=7.58$, $p<.01$; $F_2(1,41)=5.96$, $p<.05$), with the words in the high ambiguous condition being read about 15 milliseconds per word faster than the words in the low ambiguous condition This difference may to be due to a global length effect. In the low conditions, the relative clauses contain sentential complements instead of the simple direct object NPs found in the high conditions. This results in the sentences in the low condition being on average 3.5 words longer than the sentences in the high conditions. Even though the subjects did not know the content of the sentences until they read them, they were able to see how long the sentences were, and they apparently read more slowly in the longer sentences.

5.2.4.1.2 Comprehension Questions

After each trial, subjects were asked to answer a yes/no comprehension question about the sentence that had just been read. Table 2 summarizes the results for the high span subjects.

Table 2: Expt. 1 Mean Comprehension Question Scores for High Span Subjects

Experimental Conditions	% questions answered correctly
Low Ambiguous	80%
Low Unambiguous	87%
High Ambiguous	90%
High Unambiguous	87%

In the comprehension questions, there was a main effect of attachment (high or low) ($F_1(1,21)=4.57$, $p<.05$; $F_2(1,42)=3.99$, $p<.05$), as well as an interaction between attachment and ambiguity ($F_1(1,21)=3.93$, $p<.05$; $F_2(1,42)=4.26$, $p<.05$). In the low conditions, there was a main effect of ambiguity that was marginal by items and non-significant by subjects ($F_1(1,21)=2.54$, $p<.12$; $F_2(1,42)=3.62$, $p<.06$). There was no effect of ambiguity in the high conditions (all $p>.1$). From this set of statistics, it is clear that subjects had the most difficulty answering the questions for the low ambiguous condition, the same condition where they showed significant slowdowns at the disambiguating reflexive.

5.2.4.1.3 Summary of Results for High Span Subjects

Summarizing the results from the high span subjects, the effects at the disambiguating reflexive strongly suggest that the ambiguous verb was consistently parsed high (i.e. as the matrix verb) in the ambiguous conditions. The fact that the verbs in the high unambiguous condition and both of the ambiguous conditions were read more slowly than the verbs in the low unambiguous condition is consistent with high attachment of the verb in the ambiguous conditions. It is not clear whether the cost associated with high attachment of the verb is due to the search for an attachment site (as suggested above), or if making the non-local attachment itself is more difficult than making the local attachment. The finding that the high unambiguous condition was read more slowly in the early regions than the high ambiguous condition is consistent with a cost for a pronoun antecedent search.

As mentioned at the beginning of the discussion of the high-span subjects, there were no significant main effects of verb class, nor were there any significant interactions with verb class.

5.2.4.2 Low Span Subjects

Low span subjects were defined as those subjects having a reading span score of less than 2.5 (n=26).

5.2.4.2.1 Reading Times

In the discussion of the low-span subjects that follows, graphs showing both combined verb classes (as done for the high span subjects) and separated verb classes will be presented. The following graphs show the average residual reading times for the low memory-span subjects in the conditions requiring low attachment.

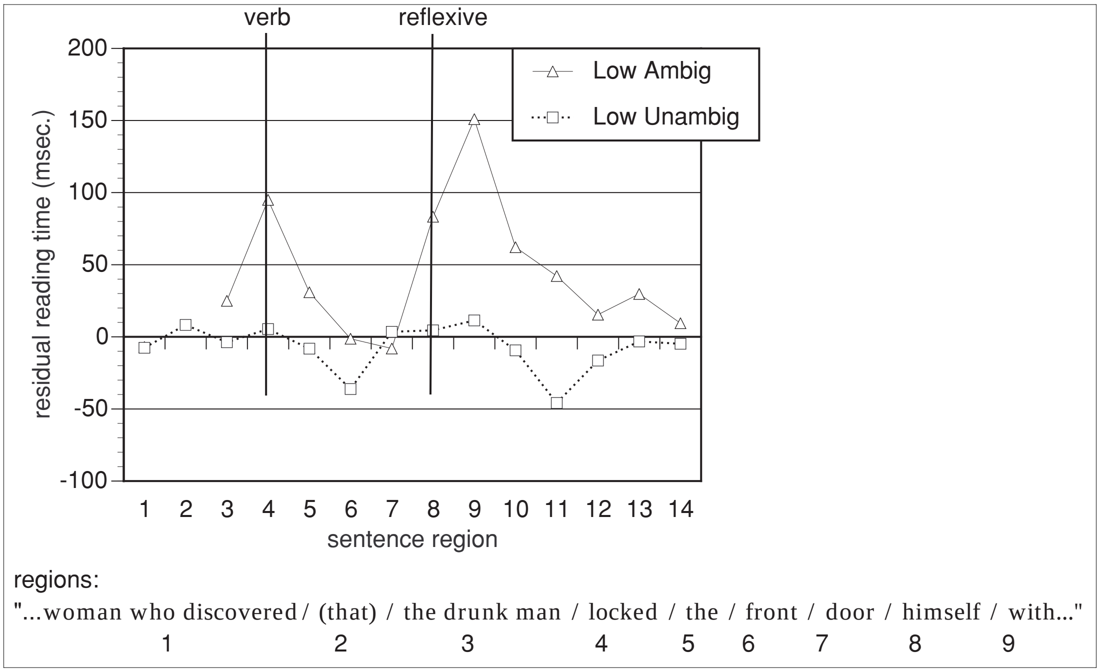


Figure 8: Expt. 1, Low Span Subjects, Low Conditions, Verb Classes Combined, n=26

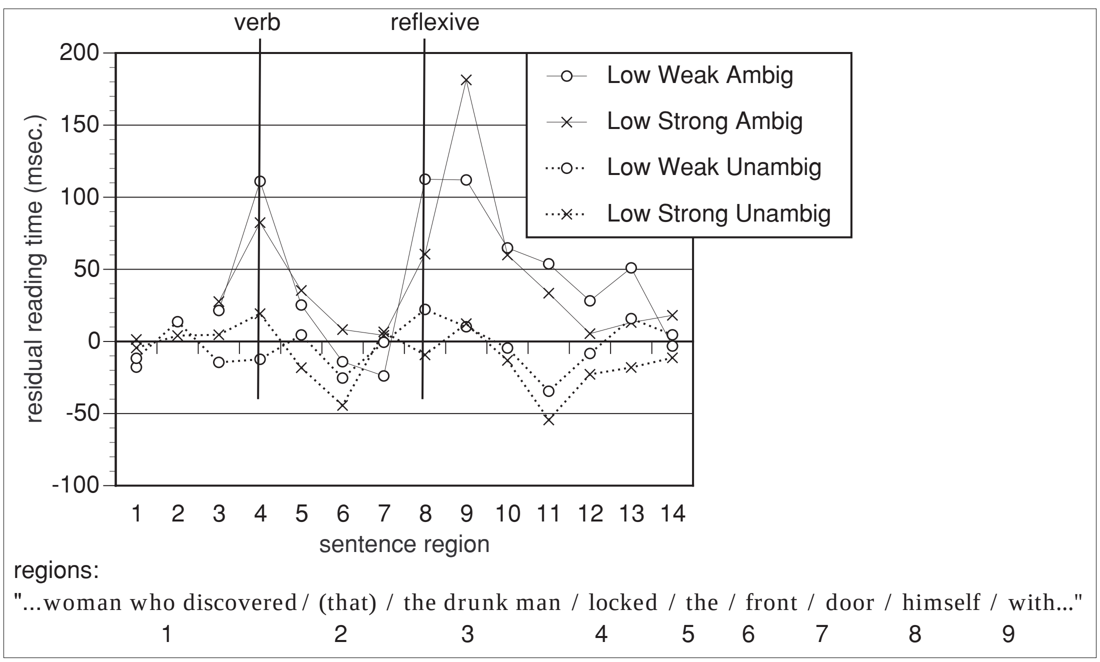


Figure 9: Expt. 1, Low Span Subjects, Low Conditions, n=26

There was a significant main effect of ambiguity at the word following the reflexive (region 9) ($F_1(1,25)=30.3$, $p<.01$, $F_2(1,41)=27.5$, $p<.01$), which suggests that the subjects did not uniformly make the low attachment in the ambiguous conditions. This effect was also seen at region 8 ($F_1(1,25)=7.58$, $p<.01$; $F_2(1,41)=8.63$, $p<.01$), region 10 ($F_1(1,25)=11.70$, $p<.01$; $F_2(1,41)=11.99$, $p<.01$), and region 11 ($F_1(1,25)=19.69$, $p<.01$; $F_2(1,41)=19.52$, $p<.01$), but it was only marginally significant at region 12 ($F_1(1,25)=3.09$, $p<.08$; $F_2(1,41)=2.21$, $p<.14$) and region 13 ($F_1(1,25)=2.87$, $p<.10$; $F_2(1,41)=3.81$, $P<.06$).

There was a significant region x ambiguity interaction at regions 7-8 ($F_1(1,25)=6.27$, $p<.05$; $F_2(1,41)=6.27$, $p<.05$), at regions 7-9 ($F_1(2,25)=9.16$, $p<.01$; $F_2(2,41)=9.23$, $p<.01$), and at regions 9-10 ($F_1(1,25)=4.62$, $p<.05$; $F_2(1,41)=4.46$, $p<.05$). This interaction provides further support for the idea that the ambiguous and unambiguous conditions involve different types of processes. Similar to the high span subjects, it appears that some high attachments were made in the ambiguous conditions, and the interaction between region and ambiguity reflects the revision necessary when the ambiguous condition is disambiguated towards a low attachment. The interaction was only marginally significant at regions 9-11 ($F_1(2,25)=2.63$, $p<.08$; $F_2(2,41)=2.59$, $p<.08$), and was not significant at regions 8-9, 8-10, 10-11, and 10-12 (all $p>.10$).

In the low conditions, there were no significant effects of verb class, nor any interactions involving verb class (all $p>.1$). As can be seen in the graphs, there were numeric differences between the verb classes at regions 8 and 9, but because the effects were in different directions in the two regions there is no reason to believe that they are of any consequence.

Figure 10 and Figure 11 show the results from the low-span subjects in the high conditions.

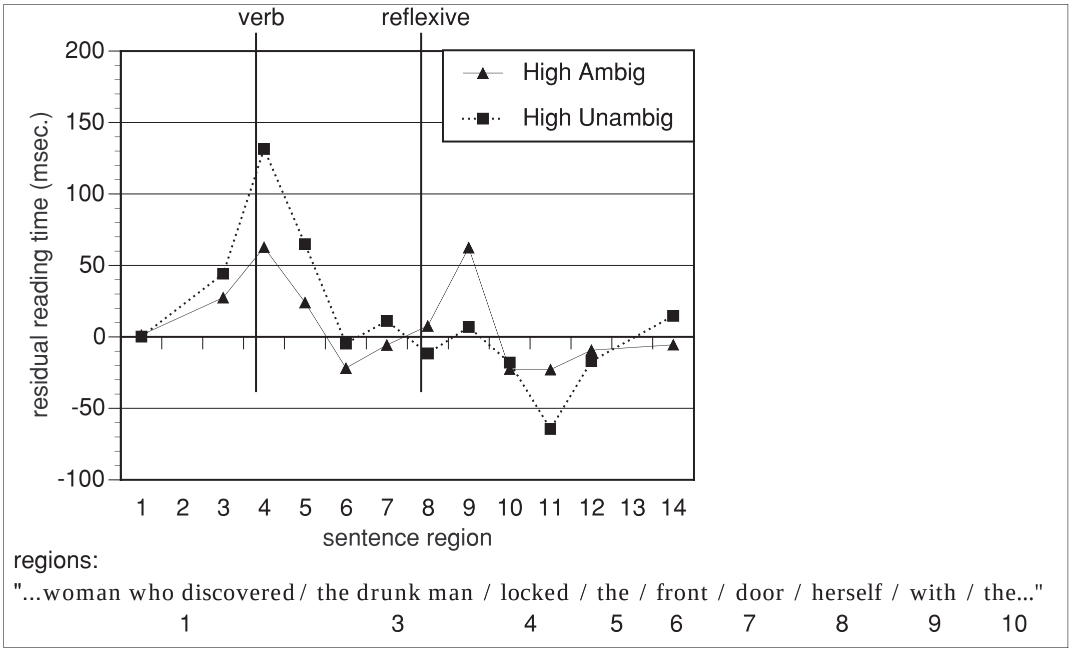


Figure 10: Expt. 1, Low Span Subjects, High Conditions, Verb Classes Combined, n=26

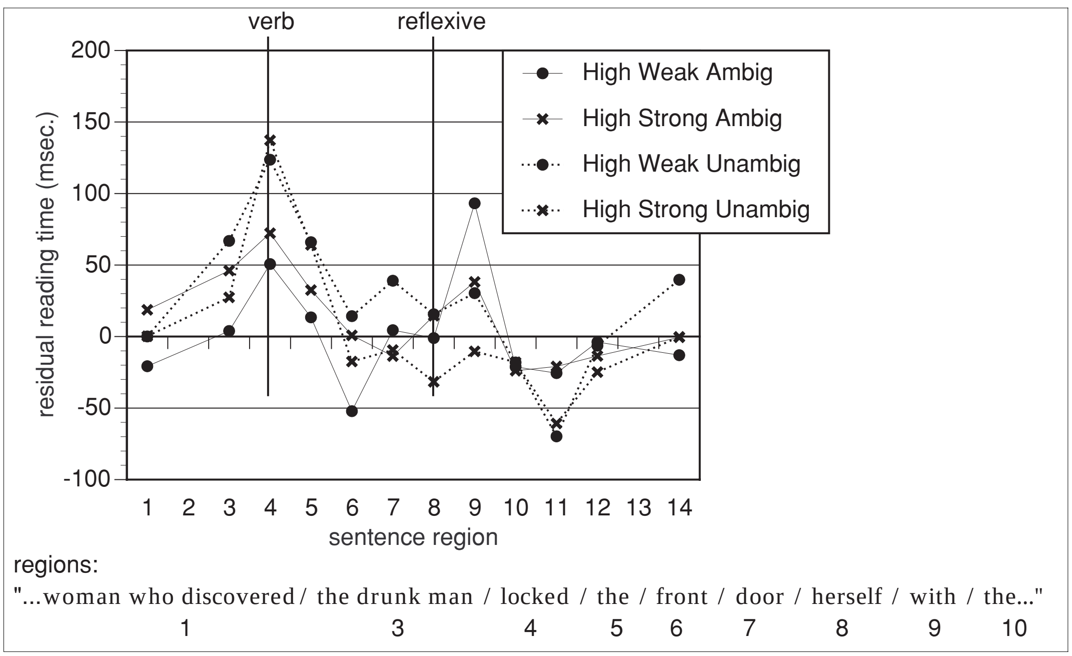


Figure 11: Expt. 1, Low Span Subjects, High Conditions, n=26

In the high conditions, there was a significant main effect of ambiguity at the word following the reflexive ($F_1(1,25)=7.79$, $p<.01$; $F_2(1,42)=7.42$, $p<.01$). This finding suggests that the low-span subjects did not uniformly make the high attachment in the ambiguous conditions. This is different from the high-span subjects, who appeared to always make the high attachment in the ambiguous conditions. The main effect of ambiguity was only found at region 9; at regions 8 and 10 it was not significant (all $Fs<1$).

Further support for the idea that the subjects sometimes made the low attachment in the ambiguous conditions is found in the significant interaction between ambiguity and region at regions 7-9 ($F_1(2,25)=3.10$, $p<.05$; $F_2(2,42)=3.15$, $p<.05$), at regions 9-10 ($F_1(1,25)=5.98$, $p<.05$, $F_2(1,42)=5.81$, $p<.05$), at regions 9-11 ($F_1(1,25)=3.57$, $p<.05$; $F_2(2,25)=3.38$, $p<.05$), and at regions 10-11 ($F_1(1,25)=4.93$, $p<.05$; $F_2(1,42)=4.37$, $p<.05$). The interaction was marginal at regions 8-10 ($F_1(2,25)=2.68$, $p<.06$; $F_2(2,42)=2.66$, $p<.08$), and was not significant at regions 7-8, 8-9, 11-12, or 10-12 (all $p>.10$).

There was also a main effect of verb class in the high conditions at region 9 which was significant by subjects, but not by items ($F_1(1,25)=6.97$, $p<.01$; $F_2(1,42)=2.71$, $p<.11$). The effect was not significant at region 8 (all $Fs<1$). There was no interaction between verb class and ambiguity at either region 8 or region 9 (all $p>.1$).

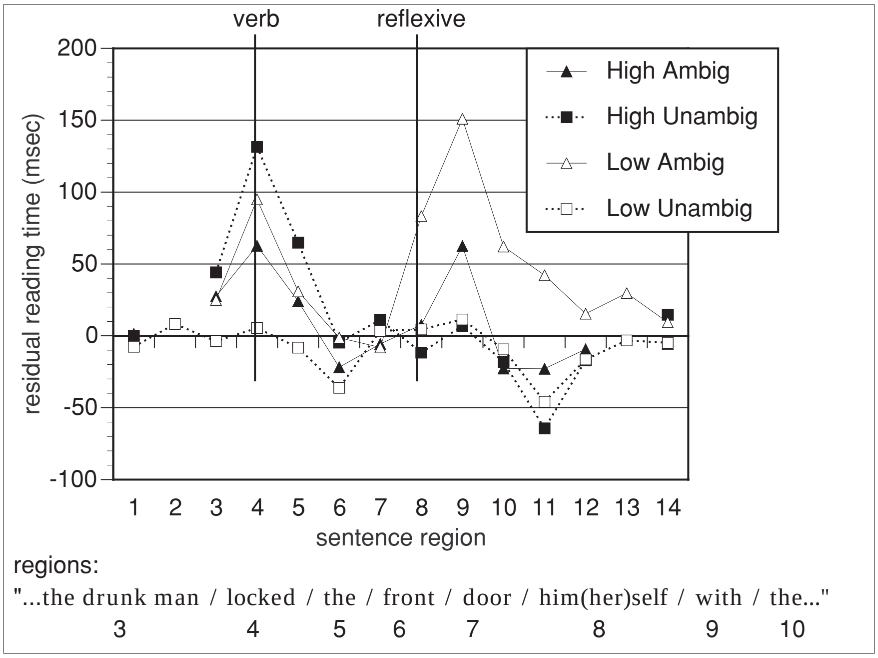


Figure 12: Expt. 1, Low Span Subjects, Verb Classes Combined, n=25

As will become clear in what follows, the low span subjects behaved similarly to the high-span subjects in the earlier, ambiguous regions. In the low conditions at the ambiguous NP (region 3), there was a significant main effect of ambiguity ($F_1(1,25)=6.10$, $p<.05$; $F_2(1,41)=7.51$, $p<.01$). In these conditions, there was no main effect of verb class, nor was there a verb class x ambiguity interaction (all Fs<1). In the high conditions, there were no effects of ambiguity or verb class (all Fs<1) at the ambiguous NP. These findings indicate that the subjects treated the low unambiguous condition different from the rest of the conditions, which would be expected if the ambiguous conditions all involve high attachment, while the low unambiguous condition involves low attachment because of the preceding complementizer *that*.

There was a significant interaction between verb class and ambiguity ($F_1(1,25)=5.38$, $p<.05$; $F_2(1,41)=6.49$, $p<.05$) at the region 3 NP in the high conditions. The difference between unambiguous and ambiguous conditions was greater in the weakly-biased condition (0 (amb.) vs. 66 (unamb.)) than in the strongly NP-biased condition (47 (amb.) vs. 23 (unamb.)), and the pattern of difficulty was reversed between the two verb classes, with the ambiguous sentences more difficult in the strongly-biased conditions and the unambiguous sentences more difficult in the weakly-biased conditions. This finding is consistent with the existence of an extra cost for unambiguous pronouns that are not strongly supported by the verb bias (or a lower cost for pronouns that are supported by the verb bias), since the weakly NP-biased condition was the one with the slowest reading time for the pronoun.

At the verb (region 4), there was a significant main effect of ambiguity in the low conditions ($F_1(1,25)=12.5$, $p<.01$; $F_2(1,41)=16.0$, $p<.01$) and in the high conditions ($F_1(1,26)=6.85$, $p<.01$, $F_2(1,41)=5.62$, $p<.05$), but there was no main effect of verb class, nor were there any interactions with verb class in either condition (all $Fs<1$). Just as for the high-span subjects, the high unambiguous condition was read most slowly, the low unambiguous condition was read most quickly, and the ambiguous conditions where in the middle.

As with the high span subjects, there was some effect of the attachment site in the ambiguous conditions, but the effect was less than for the high-span subjects. The low-span subjects read regions 4-6 of the ambiguous conditions faster in the low conditions than in the high conditions, but the difference was not significant by subjects and only marginally significant by items ($F_1(1,25)=2.05$, $p>.1$; $F_2(1,41)=3.48$, $p<.09$). For the entire identical region (regions 1-7), there was no main effect of attachment site (all $Fs<1$).

5.2.4.2.2 Comprehension by Low Span Subjects

The average scores on the comprehension questions for the low-span subjects can be found below in Table 3:

Table 3: Expt. 1 Mean Comprehension Question Scores for Low Span Subjects

Experimental Condition	% questions answered correctly
Low Ambiguous	81%
Low Unambiguous	87%
High Ambiguous	95%
High Unambiguous	90%

For the low-span subjects, there was a significant main effect of attachment site (high or low) ($F_1(1,25)=19.7$, $p<.01$; $F_2(1,42)=21.98$, $p<.01$) and an ambiguity x attachment interaction ($F_1(1,25)=7.36$, $p<.01$; $F_2(1,42)=8.03$, $p<.01$). There was no overall main effect of ambiguity (all Fs<1). When the low conditions were analyzed separately, there was a main effect of ambiguity that was significant by items and marginally significant by subjects ($F_1(1,25)=3.49$, $p<.07$; $F_2(1,42)=4.26$, $p<.05$). There was also a main effect of ambiguity in the high conditions ($F_1(1,25)=4.19$, $p<.05$; $F_2(1,42)=4.21$, $p<.05$), with the ambiguous conditions being understood better. The fact that the subjects found the low ambiguous condition more difficult than the low unambiguous condition is in accord with the finding at the disambiguation, where the subjects read the ambiguous condition more slowly than the unambiguous condition. It is not clear how to interpret the finding that the high ambiguous condition was understood more accurately than the high unambiguous condition. This data is the opposite of what would be expected from the reading times at the disambiguation, where there was a small slowdown in the ambiguous condition relative to the unambiguous condition. Note the the high-span subjects showed no significant difference in accuracy between the ambiguous and unambiguous high conditions.

5.2.4.3 Subject Group Comparison

The results presented so far have suggested that the high span subjects consistently make the high attachment in the ambiguous conditions, while the low span subjects make both the high and low attachments in the ambiguous conditions. If this is so, then there should also be interactions involving subject group. At the word immediately following the reflexive (region 9) there is no main effect of subject group (p>.1). There is, however a significant interaction between ambiguity and subject group ($F_1(1,47)=4.65$, $p<.05$, $F_2(1,42)=4.37$, $p<.05$). There are no other significant interactions with subject group ($p>.1$).

Recall that in the high conditions, the high span subjects have no trouble at the disambiguating reflexive, while the low span subjects experience a slowdown at the reflexive. This difference is reflected in a significant interaction between ambiguity and subject group ($F_1(1,47)=4.13$, $p<.05$; $F_2(1,42)=4.48$, $p<.05$). In the low conditions, where both groups of subjects experienced difficulty, there is no similar ambiguity x subject group interaction ($p>.1$). Thus, these interactions confirm that the subjects behave similarly in the low conditions, but not in the high conditions.

To summarize the results for the low span subjects, there is significant evidence of both high and low attachments in the ambiguous conditions. Additionally, there are some effects of verb class in the high conditions (at both the ambiguous verb and the disambiguating reflexive), but none in the low conditions. One possible reason for this is that verb bias information affects the ease or difficulty of processing the pronoun in the high unambiguous conditions, but doesn t affect the processing in the low control conditions (due to the strong influence of the complementizer). Verb bias could also have an effect on how difficult it is to reanalyze into the high position (i.e. how difficult it is to take the sentential complement away from the embedding verb). In the ambiguous regions (3-5), the low span subjects read most quickly in the low unambiguous condition, when it is clear that the words must be in a sentential complement. The ambiguous conditions are somewhat more difficult and the high unambiguous condition is the most difficult. These findings are consistent with a cost to search for the high attachment for the verb, and with a cost to search for pronoun antecedents.

5.3 Experiment 2

A second experiment was performed to test several questions left open by the first experiment. First, the results of the first experiment could be due to either a general preference for high attachment, or they could be the result of following frequency biases (recall that all of the ambiguous verbs in Experiment 2 were either weakly or strongly biased towards NP complements). In an effort to distinguish these alternatives and to broaden the generality of the results, a broader range of verbs was tested, this time including a class of verbs biased towards sentential complements and a class of verbs which allow only direct object complements.

Second, since the method of early disambiguation differed between the high and low control conditions in Experiment 1, it is not possible to directly compare the early regions of the control sentences. In the second experiment, the disambiguator in the low unambiguous conditions was changed from a complementizer to a nominative case-marked pronoun, making it comparable to the accusative pronoun that disambiguates the high unambiguous conditions.

A third possible problem with Experiment 1 has to do with the conditions that were disambiguated with a case-marked pronoun. In the first experiment these pronouns were introduced without any antecedent, which makes the sentences rather unnatural. Studies of pronoun resolution (Clark and Sengul 1979, Fischer and Glanzer 1986) have shown that the farther away the antecedent is, the more slowly the pronoun will be processed. The fact that there was no antecedent at all in Experiment 1 may mean that the pronoun search takes longer just because there is no referent. To deal with this concern, the ambiguities in the second experiment were embedded within an additional clause containing an antecedent for the pronoun. While this made the stimuli longer and slightly harder to process, accuracy levels fell only slightly, about 6%, from the accuracy level in Experiment 1. Another factor likely contributing to the lower scores is that the comprehension questions in Experiment 2 all questioned the ambiguity, thereby making them on average more difficult than the questions in Experiment 1, where the ambiguity was not always questioned.

5.3.1 Materials

Experiment 2 included the same four conditions (high and low attachment, ambiguous and unambiguous) that were in Experiment 1. An example of all four conditions for one stimulus set is presented below:

(220) *Low Ambiguous*
The talent agency thinks that the creative woman who knows the funny man wrote some comedy sketches himself about the amusing escapades wants to publish them.

(221) *Low Unambiguous*
The funny man thinks that the creative woman who knows him wrote some comedy sketches herself about the amusing escapades she had seen.

(222) *High Ambiguous*
The talent agency thinks that the creative woman who knows the funny man wrote some comedy sketches herself about the amusing escapades she had seen.

(223) *High Unambiguous*
The funny man thinks that the creative woman who knows him wrote some comedy sketches herself about the amusing escapades she had seen.

The items consisted of a matrix clause whose verb unambiguously takes a tensed sentential complement. The subject of the sentential complement is modified by a subject relative clause. The lexically ambiguous verb (*knows*) is the verb of the relative clause. This verb will again be called the embedding verb . Following the embedding verb was a one-word (unambiguous conditions) or three-word (ambiguous conditions) NP, which was in turn followed by a verb (the ambiguous verb) that could take as its subject either the lowest NP (the low NP) or the subject of the highest embedded clause (the high NP). The high NP and the low NP were human NPs of opposite gender. A gender-marked emphatic reflexive was four words beyond the ambiguous verb to disambiguate towards either a high or low attachment. Immediately following the reflexive was a four-word PP. Additional material followed the PP to make the sentences more natural and to ensure that the measurements from the PP were not influenced by sentence-final wrap-up effects. In the low conditions, this extra material included the verb for the higher embedded clause that was necessary to make the sentence grammatical.

Experiment 2 included 80 sets of experimental items, 20 from each of the following classes of verbs: NP-only, Strong NP-bias, Weak NP-bias, S-bias. Each subject saw 10 items from the NP-only condition, and 20 from each of the other conditions. This resulted in the subjects each seeing 5 items from the high ambiguous conditions and 5 from the high unambiguous conditions of each verb class (the NP-only verbs had no low conditions, since that would have been ungrammatical). The strongly and weakly NP-biased stimuli were modified versions of the stimuli in the first experiment, with the exception that the sentences with *fear* were not used in Experiment 2. The S-bias, strong NP-bias, and weak NP-bias verbs used in the experiment all allowed both NP-complements and sentential complements. Additionally, the embedding verbs all allowed animate NPs as direct objects. Table 4 shows the verbs used in the experiment, along with the data used to determine the bias classification.

Table 4: Sentence Completion Data for Verbs in Experiment 2
Data for strongly and weakly NP-biased sentences repeated from Table 1 above.

	NP completions	***that*-less S-complements**	**% S-complements**
Neutral Bias			
know	31	29	48%
mention	41	17	29%
doubt	42	20	32%
notice	58	16	22%
NP Bias			
hear	74	3	4%
discover	69	9	12%
acknowledge	69	5	7%
appreciate	95	0	0%
warn	74	2	3%
understand	72	0	0%
S-Bias			
claim	6	28	82%
believe	14	20	59%
suspect	18	26	59%

The data come from the same completion studies (Susan Garnsey, p.c.) that were discussed with Experiment 1. The final column shows the percentage of these ambiguous completions that were completed with sentential complements. In addition to the *that*-less sentential complements, there were also many S-complement completions that included the complementizer *that*. These completions were excluded from the calculations used to determine verb class, since the goal of this experiment was to find out what subjects do in the ambiguous conditions. Including the S-complements with *that* would have resulted in the same groupings of verbs, though with higher percentages of S-complements (generally 15%-20% higher).

In addition, the following NP-only verbs were used: *abuse, admire*, adore, annoy, defy, despise*, disappoint, feed, fire, harass, humiliate, love*, overcharge, pity, prosecute, rescue, supervise, support, treat.* The verbs marked with an asterisk can arguably take a sentential complement, but informants judged that they were ungrammatical without a complementizer, as can be seen in (224)-(226).

(224) *I despise he eats my lunch out of the refrigerator every day.

(225) *I admire the professor can speak calmly before so many hostile students.

(226) *The woman loves her boyfriend brings her flowers every week.

Thus, like the verbs without asterisks, the ambiguous conditions for these verbs were actually unambiguous. The items in the NP-only condition contained only the high ambiguous and unambiguous conditions, since the low attachments would have been ungrammatical. As noted earlier, the subjects were only shown 10 items from this class so that they would see five items per condition (high ambiguous and high unambiguous), just as with the other verb classes

The region encodings were the same as in the first experiment. Because there is no complementizer to disambiguate the low unambiguous condition in Experiment 2, there is no region 2 (the next region was called region 3 to make region encoding equivalent across experiments).

(227) The talent agency thinks that the creative woman who knows / the funny man /
1 / 3 /
wrote / some / comedy / sketches / **himself** / about / the / amusing / escapades /
4 / 5 / 6 / 7 / 8 / 9 / 10 / 11 / 12 /
wants / to publish them.
13 / 14

130 filler sentences were also included in the experiment. 84 of the fillers were experimental sentences from two unrelated experiments that were run concurrently; the other 46 were distracters.

5.3.2 Method

The materials were presented to subjects on a computer screen using the word-by-word self-paced moving-window paradigm (Just et al. 1982), just as in Experiment 1. The stimuli were split into two blocks of 100 items, and the reading-span test was run during the break between the two blocks. The order of the two blocks was counter-balanced across subjects, and the sentences were randomized by subject within the blocks. All of the comprehension questions questioned the ambiguity. All other aspects of the method were the same as in Experiment 1.

As with for Experiment 1, all times reported here are mean adjusted residual reading times of the trials for which the subject answered the comprehension question correctly. Residual reading times longer than 8 seconds were removed completely from the analysis (this affected 3 words), and reading times that were more than 4 standard deviations from the mean were trimmed to 4 s.d. (1194 ms.). This effected less than 1% of the words (909 out of 101197 total words).

5.3.3 Subjects

64 undergraduates from the University of Delaware participated in the study. All were paid for their participation. Subjects were eliminated from the analysis if they scored less than 80% on the comprehension questions on the 46 distracters, or if they scored less than 65% on the experimental questions. 8 subjects were eliminated from the analysis due to low comprehension scores, leaving 56 subjects in the analysis.

5.3.4 Results

In the discussion of the results that follows, the subjects will again be split by reading span. As in Experiment 1, the subjects were divided into two groups as close to equal size as possible. In this experiment, subjects with a reading span score of less than 2.3 were put in the low-span group, while those with scores greater than 2.3 were put in the high-span group. This grouping yielded a high span group with 29 subjects and a low span group with 27 subjects. Because the average reading span scores in Experiment 2 were somewhat lower than in Experiment 1, the threshold was set at 2.3 in Experiment 2, instead of the 2.5 used in Experiment 1. Eight subjects had a reading span score of 2.4, so using the same threshold as in Experiment 1 would have yielded a low span group with 35 subjects and a high span group with only 21 subjects.

For each condition, the reading times at the disambiguating region will be discussed first, to establish how the subjects actually attached the ambiguous verb. The earlier regions will then be discussed in an effort to illuminate the details of the parsing process. The analyses performed were repeated measures ANOVAs with region, ambiguity, verb class, and attachment site as within subject factors.

Item 4 in Block B was removed from the analysis because the PP following the reflexive was missing the adjective, thereby making it lack a region 10. As a result, the final analysis includes 20 items in each verb class except for the weakly NP-biased group, which contained 19 items.

5.3.4.1 High Span Subjects

High span readers were defined in this experiment as those having a reading-span test score of greater than 2.3 (n=29).

5.3.4.1.1 Reading Times

The following graphs show the average residual reading times of the high memory-span subjects.

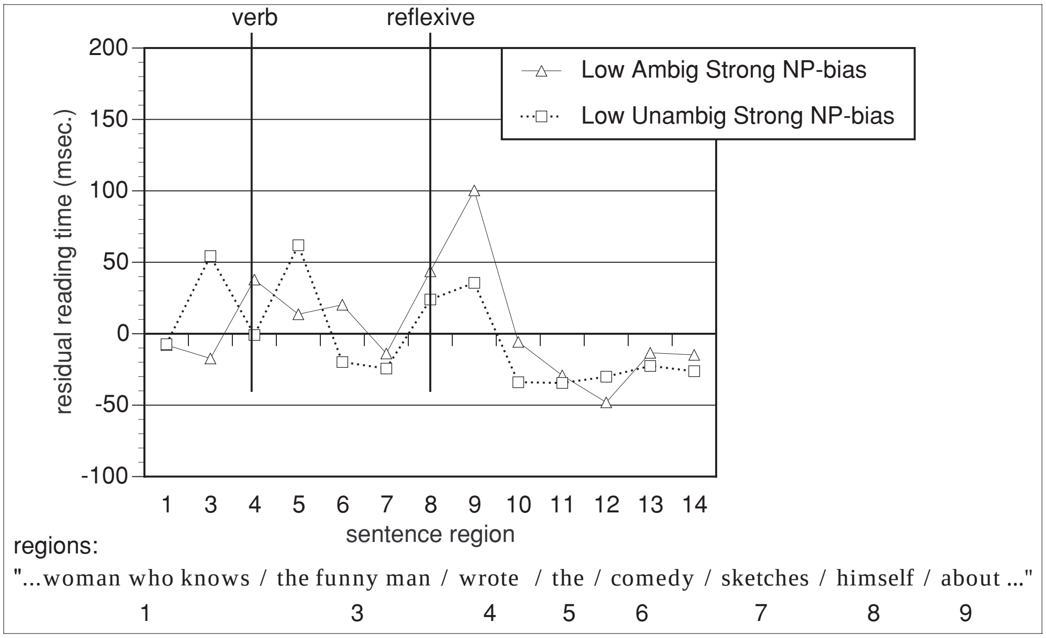


Figure 13: Expt. 2, High Span Subjects, Low Strong NP-bias Conditions, n=29

In the strongly NP-biased conditions, there was a main effect of ambiguity at the word immediately following the reflexive (region 9) which was significant by subjects, but only marginally significant by items ($F_1(1,28)=5.64$, $p<.05$; $F_2(1,19)=3.41$, $p<.07$). This indicates that the subjects made high attachments in some portion of the ambiguous trials.

At the beginning of the ambiguity (the NP in region 3), there was a significant effect of ambiguity ($F_1(1,28)=7.80$, $p<.01$; $F_2(1,19)=12.4$, $p<.01$) with the unambiguous condition read more slowly than the ambiguous condition. In Experiment 1 there were no effects of ambiguity at region 3 in any of the low conditions, although effects were found in the high conditions. In Experiment 1, the low controls were disambiguated with a complementizer *that*, while in Experiment 2 they were disambiguated with a nominative pronoun. If the effect at region 3 is an effect of processing a pronoun, the same increase should be seen for the pronoun in both the high and low ambiguous conditions in this experiment. Possible reasons for a slowdown at the pronoun include a search for a pronoun antecedent (as suggested in the discussion of Experiment 1), a length effect not factored out by the regression equation, an effect created by the regression, or an effect of the fact that the pronoun violated the expectations created by the embedding verb.

At the ambiguous verb (region 4), there was no effect of ambiguity (all Fs<1). Although there was no ambiguity effect at region 4, there was an ambiguity x region interaction between regions 3 and 4 ($F_1(1,28)=7.88$, $p<.01$; $F_2(1,19)=7.98$, $p<.01$). This indicates that the ambiguous condition became much more difficult relative to the unambiguous condition in region 4 than it was at region 3. Surprisingly, at region 5, the word following the verb, there was a main effect of ambiguity, which was marginally significant by items, though not significant by subjects ($F_1(1,28)=1.62$, $p<.21$; $F_2(1,19)=3.20$, $p<.08$), with the unambiguous conditions again slower than the ambiguous conditions. It is unclear why the unambiguous conditions should again be slower in this region, unless it is a delayed effect of the pronoun.

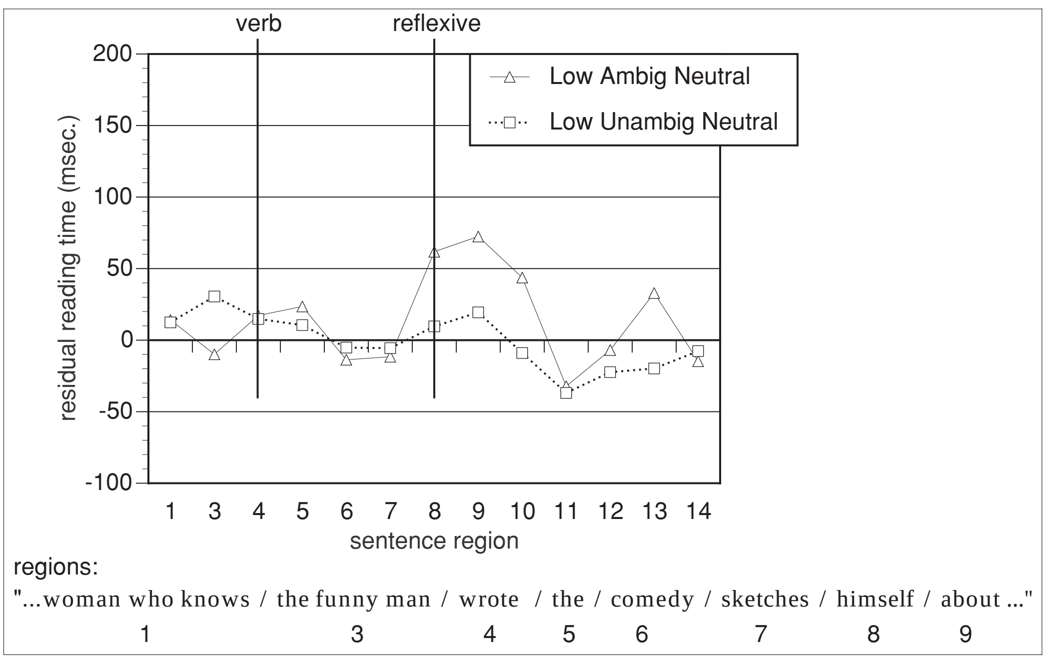


Figure 14: Expt. 2, High Span Subjects, Low Weak NP-bias Conditions, n=29

In the weakly NP-biased low conditions, there was no main effect of ambiguity at the reflexive (all Fs<1) or at the word following the reflexive ($F_1(1,28)=1.84$, $p<.18$; $F_2(1,18)=2.43$, $p<.13$), despite the large numerical difference (52 msec. difference, ambig: 72, unambig: 19). However, there was a marginally significant main effect of ambiguity two words after the reflexive (region 10) ($F_1(1,28)=3.79$, $p<.06$; $F_2(1,18)=3.67$, $p<.06$). This main effect of ambiguity was also marginally significant when regions 8 and 9 were combined ($F_1(1,28)=2.92$, $p<.09$; $F_2(1,18)=3.21$, $p<.08$), and it became significant when regions 9 and 10 were combined ($F_1(1,28)=5.38$, $p<.05$; $F_2(1,18)=5.80$, $p<.05$) and when region 8,9, and 10 were combined ($F_1(1,28)=5.93$, $p<.05$; $F_2(1,18)=6.24$, $p<.01$). This effect of ambiguity suggests that, like the strongly NP-biased conditions, (some) high attachments were made in the ambiguous condition.

At the NP in region 3, there was a main effect of ambiguity ($F_1(1,28)=4.62$, $p<.05$; $F_2(1,18)=3.78$, $p<.06$), with the unambiguous condition (the pronoun) again read more slowly than the ambiguous full NP. There was no main effect of ambiguity at the verb (all Fs<1), nor was there a region x ambiguity interaction between regions 3 and 4 (all p>.1).

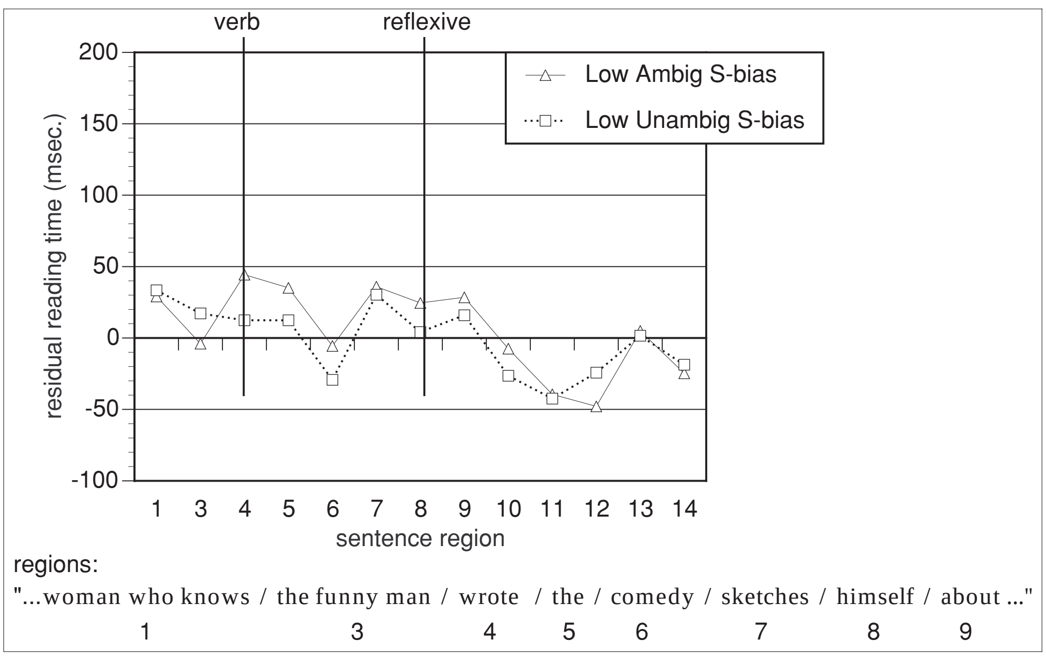


Figure 15: Expt. 2, High Span Subjects, Low S-bias Conditions, n=29

In the S-bias low condition, there was no main effect of ambiguity at the word following the reflexive (all Fs<1), nor was there an effect at the reflexive itself (all Fs<1). This lack of effect suggests that the subjects always attached the verb low in the ambiguous conditions. There was no main effect of ambiguity at either the NP (region

3), at the verb itself, or at the verb and the following word (all p>.1). There was no significant region x ambiguity interaction at regions 3 and 4 (all p>.1). However, there was a marginal main effect of ambiguity over the entire ambiguous region (region 4-7) ($F_1(1,28)=2.76$, $p<.10$; $F_2(1,19)=2.22$, $p>.1$), with the ambiguous condition read more slowly than the unambiguous condition. There was no effect of ambiguity for the entire sentence (all Fs<1). This suggests that even though the subjects consistently made the low attachment, there was still some knowledge of the temporary ambiguity of the sentence.

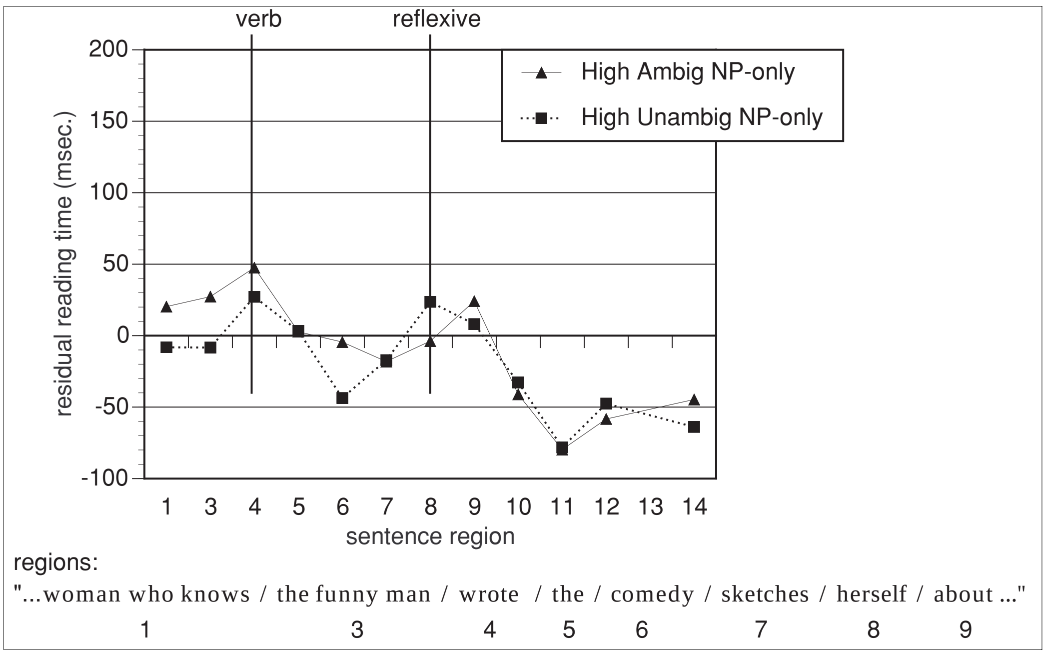


Figure 16: Expt. 2, High Span Subjects, High NP-only Conditions, n=29

In the high NP-only conditions, there was no effect of ambiguity at the reflexive or at the following word (all p>.1), suggesting that the subjects always made the high attachment in the ambiguous conditions. There was a marginal main effect of ambiguity at the NP in region 3 ($F_1(1,28)=3.04$, $p<.09$; $F_2(1,19)=1.75$, $p>.1$). There was no main effect of ambiguity at the verb (all Fs<1).

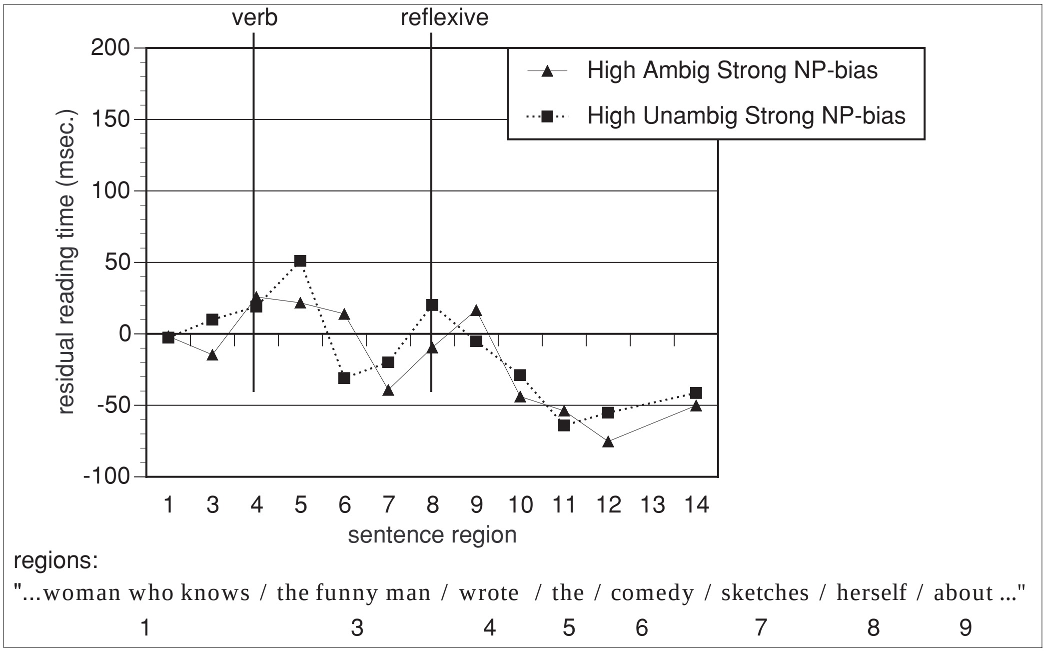


Figure 17: Expt. 2, High Span Subjects, High Strong NP-bias Conditions, n=29

The strongly NP-biased high conditions were very similar to the NP-only conditions, with no significant main effect of ambiguity at the word following the reflexive (all $p>.1$). There was no significant main effect of ambiguity at the NP in region 3 (all $p>.1$), at the verb (all $Fs<1$) or at the word following the verb (all $p>.1$).

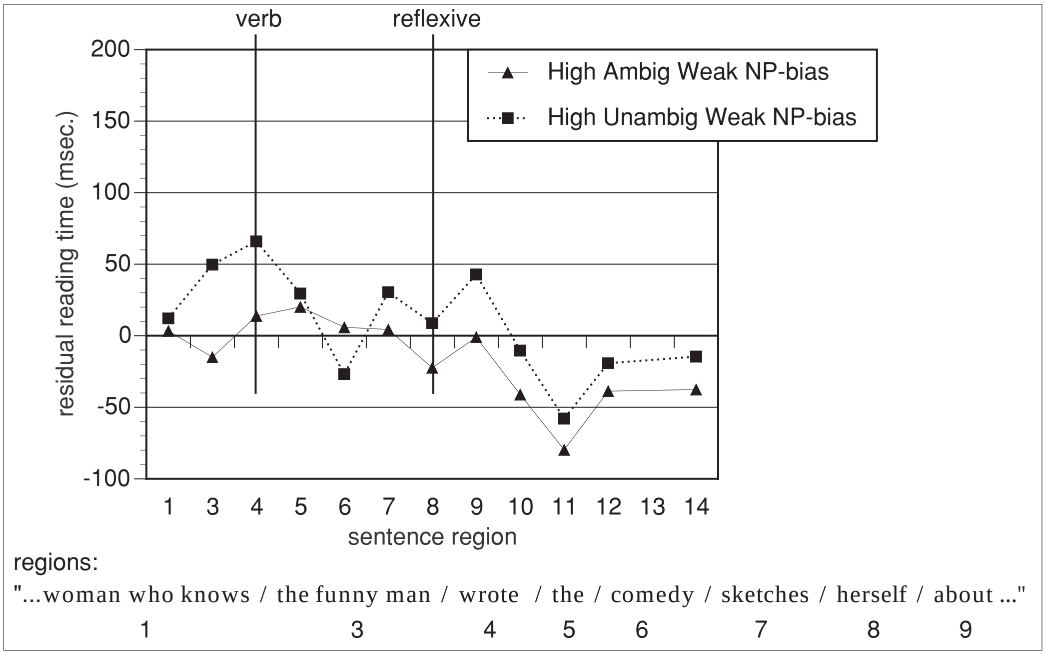
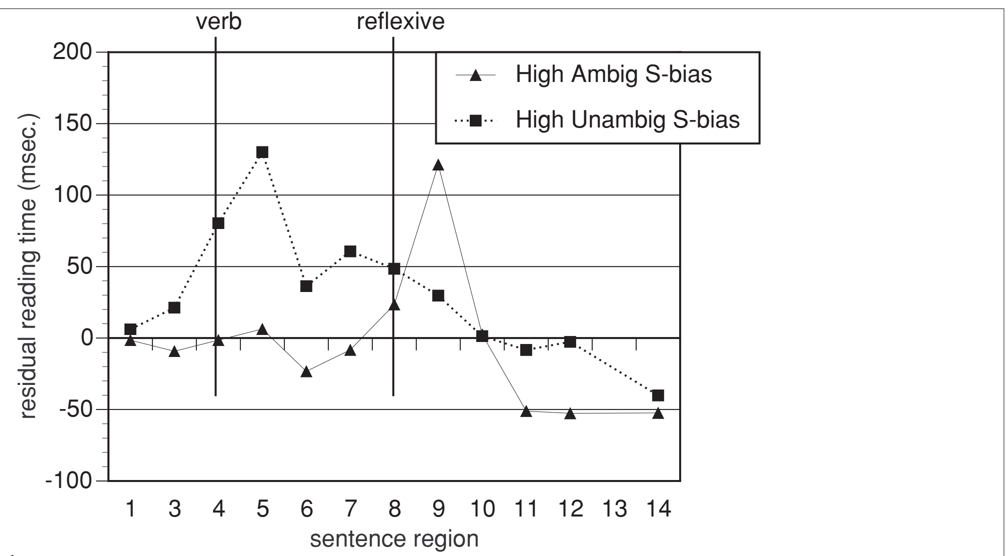


Figure 18: Expt. 2, High Span Subjects, High Weak NP-bias Conditions, n=29

The weakly NP-biased high conditions were somewhat different from the strongly-biased conditions, in that they showed a marginal effect of ambiguity at the word following the reflexive in the item analysis, but not in the subject analysis ($F_1(1,28)=1.40$, $p>.1$; $F_2(1,18)=3.83$, $p<.06$). Surprisingly, however, the unambiguous condition was read more slowly than the ambiguous condition. This effect is unexpected, given that the ambiguous condition was read more slowly at region 9 in all other conditions that showed ambiguity effects. It is not clear that the ambiguity effect in this condition is due to a processing difference caused by to the disambiguation, since there was no interaction between ambiguity and region in regions 7-9 (all Fs<1), and there was in fact a sizeable (26 msec.), though non-significant (all p>.1) main effect of ambiguity in region 7, *before* the reflexive, where the unambiguous condition was also slower then the ambiguous condition. If the ambiguity effect were due to the disambiguation, there should have been an ambiguity x region interaction when the regions immediately before and after the disambiguation were compared. In all other cases where there was an ambiguity effect at the disambiguation, it has been assumed that it is because of the effects of the disambiguation itself. The lack of an interaction (involving the regions before and after the disambiguation) indicates that the main effect is due to a general effect of ambiguity rather than a difference that occurs due to the disambiguation itself. Thus, there is no substantial evidence that the subjects did not make the high attachment in the weakly NP-biased ambiguous conditions.

At the NP in region 3, there was a significant effect of ambiguity ($F_1(1,28)=5.51$, $p<.05$; $F_2(1,18)=8.81$, $p<.01$), just as was seen in the low weakly NP-biased conditions, but different from the two high conditions just discussed. The pronoun in the unambiguous condition was again read more slowly than the full NP in the ambiguous condition. There was no main effect of ambiguity at the verb itself or at the verb and the following word (all p>.1).



regions:
"...woman who knows / the funny man / wrote / the / comedy / sketches / herself / about ..."
1 3 4 5 6 7 8 9

Figure 19: Expt. 2, High Span Subjects, High S-bias Conditions, n=29

In the S-bias high conditions, there was a significant main effect of ambiguity at the word following the reflexive ($F_1(1,28)=5.47$, $p<.05$; $F_2(1,19)=7.33$, $p<.01$), which indicates that the low attachment was made at least sometimes in the ambiguous condition. There was no main effect of ambiguity at the NP in region 3 (all p>.1), although there was a numerical tendency (30 msec. difference) for the unambiguous condition to be read more slowly. This numerical tendency towards slower reading in the unambiguous conditions is similar to the significant effects found in most of the other conditions. At the verb, there was a marginal main effect of ambiguity ($F_1(1,28)=3.34$, $p<.07$; $F_2(1,19)=3.46$, $p<.07$), with the unambiguous condition read more slowly. This main effect became significant at the word following the verb ($F_1(1,28)=9.33$, $p<.01$; $F_2(1,19)=13.41$, $p<.01$), and was even stronger when regions 4-5 were combined ($F_1(1,28)=11.83$, $p<.01$; $F_2(1,19)=14.82$, $p<.001$).

5.3.4.1.2 Comprehension Questions

Table 5 shows mean scores for the comprehension questions for the high span readers. As can be seen, the subjects were generally more accurate in the low S-bias and high NP-bias conditions, the conditions in which the verb-bias was respected and the cost of ambiguity at region 9 was lowest.

Table 5: Expt. 2 Mean Comprehension Question Scores for High Span Readers

	Ambiguous	**Unambiguous**
Low Strong NP-bias	71%	80%
Low Weak NP-bias	74%	77%
Low S-bias	83%	85%
High NP-only	92%	84%
High Strong NP-bias	81%	87%
High Weak NP-bias	82%	89%
High S-bias	72%	74%

In the low conditions, there was a significant main effect of verb class by subjects, but not by items ($F_1(2,28)=4.12$, $p<.05$; $F_2(2,56)=2.20$, $p>.1$) and a marginally significant effect of ambiguity ($F_1(1,28)=2.89$, $p<.09$; $F_2(1,56)=2.88$, $p<.09$). There were no significant interactions between verb class and ambiguity in the low conditions (all $p>.1$).

In the high conditions, there was a main effect of verb class, both when the NP-only class was included ($F_1(3,28)=9.25$, $p<.01$; $F_2(3,75)=4.31$, $p<.01$) and when the NP-only class was excluded ($F_1(2,28)=8.69$, $p<.01$; $F_2(2,56)=4.03$, $p<.05$). When the NP-only class was excluded, there was also a marginal main effect of ambiguity ($F_1(1,28)=2.95$, $p<.09$; $F_2(1,56)=3.27$, $p<.08$). This effect of ambiguity was not present when the NP-only class was included in the calculation (all Fs<1); this is not surprising, since there is no actual ambiguity in the NP-only conditions. There were no significant interactions between verb class and ambiguity in the high conditions (including or excluding the NP-only class) (all p>.1).

When all conditions were combined, the interaction between verb class and attachment site was significant ($F_1(3,28)=12.85$, $p<.001$; $F_2(3,75)=11.81$, $p<.001$), and remained significant when the NP-only conditions were excluded ($F_1(2,28)=12.16$, $p<.001$; $F_2(2,56)=11.21$, $p<.001$). This indicates the verb bias affects comprehension differently in the high and low conditions (i.e. S-bias makes the low conditions easier, while NP-bias makes the high conditions easier).

5.3.4.1.3 Verb Class Comparison

In the low conditions for the high span subjects, there was no ambiguity x verb class interaction in the reading times of the word following the reflexive (region 9) (all p>.1). If the fact there was no interaction at the disambiguation point means that ambiguity has the same effect regardless of verb bias, it would not be surprising if there were no interaction between ambiguity and verb class at the verb, where the ambiguity is initially encountered. In line with this expectation, there was no ambiguity x verb class interaction at the verb or at the verb and the following word (all Fs<1). Recall that an ambiguity effect was found when regions 4-5 were combined in the strongly NP-biased conditions. Thus, if there were any interaction, it should have been found when regions 4-5 were combined. At the NP in region 3 there was no interaction between verb class and ambiguity (all p>.1). These findings indicate that the different verb classes are treated similarly, i.e. that being forced to the low attachment is no more difficult in the strongly NP-biased condition than in the S-biased condition. This is at odds with the findings found in analyses of the individual verb classes, where there was an effect of ambiguity around the reflexive in both of the NP-biased conditions, but there was no such effect in the S-biased condition. One possible explanation is that the statistical tests performed for the interaction are not sensitive enough to find differences, even though they do appear to exist in the individual analyses.

In the high conditions, there was an ambiguity x verb class interaction at the word following the reflexive (region 9) ($F_1(2,28)=5.50$, $p<.01$; $F_2(2,56)=7.07$, $p<.01$). This confirms that the verb bias does have an effect in the NP-biased conditions there was no ambiguity effect (i.e. all high attachments), while in the S-biased condition there was an ambiguity effect (i.e. at least some low attachments). There was no interaction between verb class and ambiguity at the ambiguous verb in region 4 (all $p>.1$). However, when the verb and the following word were combined, the interaction between ambiguity and verb class was significant ($F_1(2,28)=3.88$, $p<.05$; $F_2(2,56)=3.98$, $p<.05$). This fact, similar to the interaction at region 9, suggests that the processing of the ambiguous verb is based on the properties of the embedding verb. At the region 3 NP in the high conditions, there was no interaction between ambiguity and verb class (all $p>.1$). For all of the verb class interactions, the NP-only condition was excluded, since the there is no truly ambiguous NP-only condition. If the NP-only conditions are included, the interactions remain very similar.

For the high span subjects, there was a significant 3-way ambiguity x attachment site x verb class interaction at the word following the reflexive ($F_1(3,28)=5.36$, $p<.01$; $F_2(3,75)=6.24$, $p<.01$). The interaction did not reach significance at either the reflexive (region 8) or the word two words after the reflexive (region 10) (all $p>.1$). This finding confirms that the cost of ambiguity is predicted by the degree of consistency between the verb bias and the required parse. For example, if the verb bias and attachment site agree (as in NP-bias high attachments and S-bias low attachments), there is relatively little effect of ambiguity. On the other hand, when the attachment site is not the one predicted by the verb (e.g. NP-bias low attachments and S-bias high attachments), there is a much greater effect of ambiguity.

5.3.4.2 Low Span Subjects

The low span subjects were defined as those having a reading-span score of less than 2.3 (n=27).

5.3.4.2.1 Reading Times

Below are the results for the low memory-span subjects in Experiment 2.

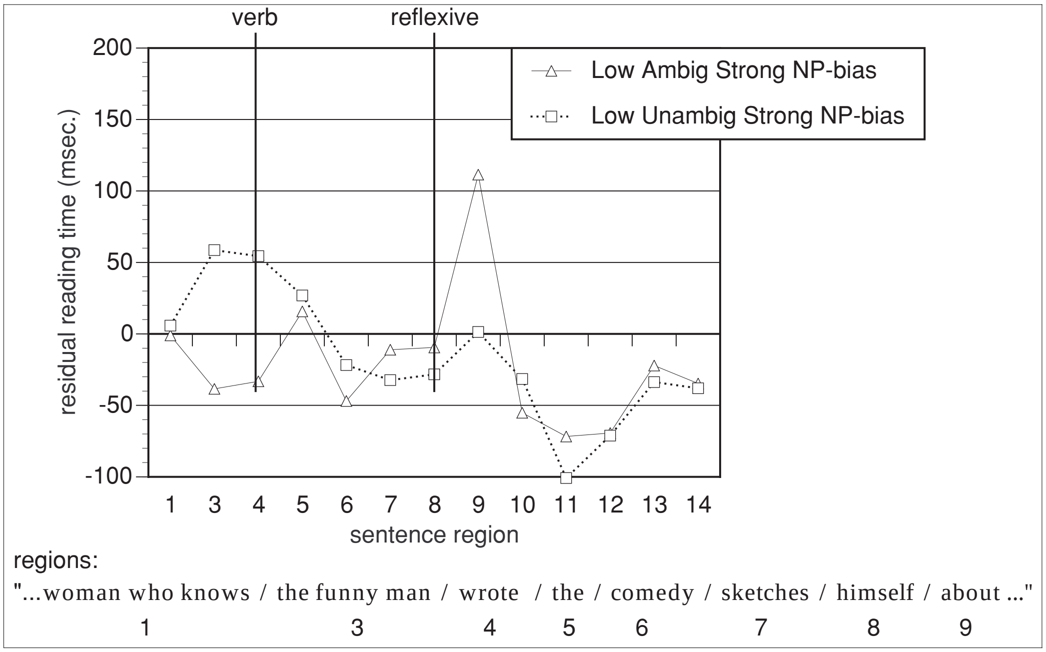


Figure 20: Expt. 2, Low Span Subjects, Low Strong NP-bias Conditions, n=27

In the low span subjects, there was a main effect of ambiguity at the word following the reflexive in the strongly NP-biased condition. ($F_1(1,26)=8.60$, $p<.01$; $F_2(1,19)=8.05$, $p<.01$). This suggests that the subjects made high attachments in at least some of the ambiguous conditions.

There was a significant main effect of ambiguity at the NP in region 3 ($F_1(1,26)=12.05$, $p<.01$; $F_2(1,19)=13.01$, $p<.01$), where the unambiguous condition was read more slowly than the ambiguous condition. There was also a main effect of ambiguity at the verb ($F_1(1,28)=4.53$, $p<.05$; $F_2(1,19)=5.67$, $p<.05$), where the unambiguous condition was again read more slowly than the ambiguous condition. There were no other significant interactions or main effects.

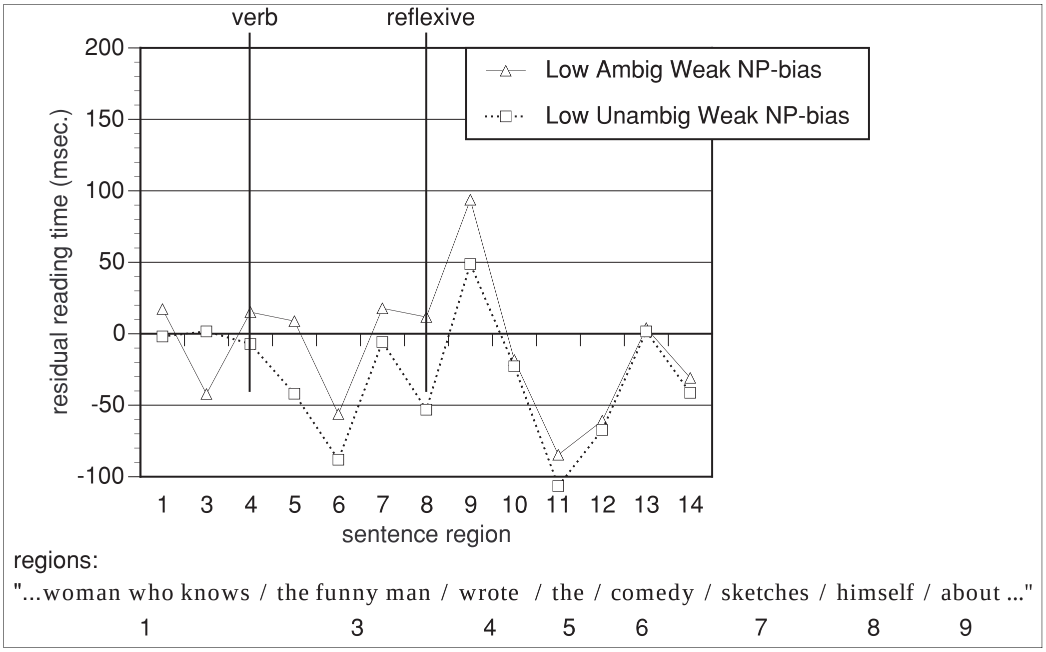


Figure 21: Expt. 2, Low Span Subjects, Low Weak NP-bias Conditions, n=27

In the weakly NP-biased conditions, there was no main effect of ambiguity at the word following the reflexive (all p>.1). When the reflexive and the following word were combined (R8-9), there was a main effect of ambiguity, which was significant by subjects, and marginally significant by items ($F_1(1,26)=4.03$, $p<.05$; $F_2(1,18)=3.25$, $p<.08$). This indicates that the subjects made at least some high attachments in the ambiguous conditions.

At the NP in region 3, there was a significant main effect of ambiguity in the subject analysis, and a marginal effect in the item analysis ($F_1(1,26)=4.58$, $p<.05$; $F_2(1,18)=2.74$, $p<.10$). There was no effect of ambiguity at the verb or at the verb and the following word combined (all p>.1), although the effect of ambiguity was significant by item at region 5 ($F_1(1,28)=2.39$, $p>.1$; $F_2(1,18)=4.02$, $p<.05$).

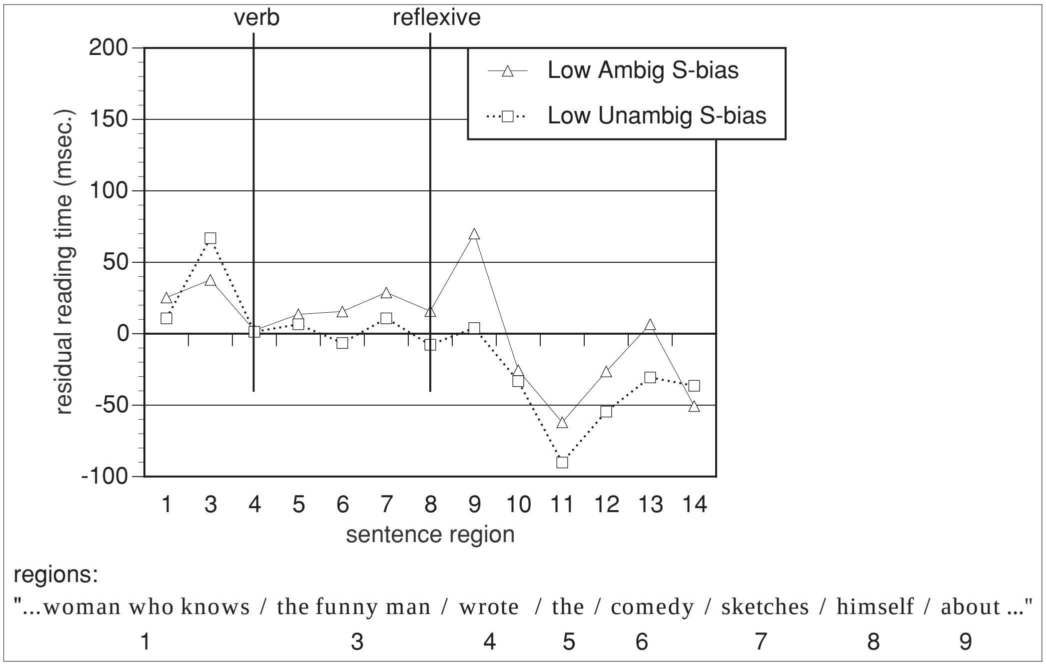


Figure 22: Expt. 2, Low Span Subjects, Low S-bias Conditions, n=27

In the low S-biased conditions, there was a marginally significant effect of ambiguity at the word following the reflexive ($F_1(1,26)=2.98$, $p<.09$; $F_2(1,19)=3.25$, $p<.08$). This suggests that the low span subjects made some high attachments in the S-biased condition, in contrast to the high span subjects, who showed no evidence for high attachments in the S-biased conditions. At the verb, there was no main effect of ambiguity (all Fs<1), nor was there an effect at the NP in region 3 (all p>.1).

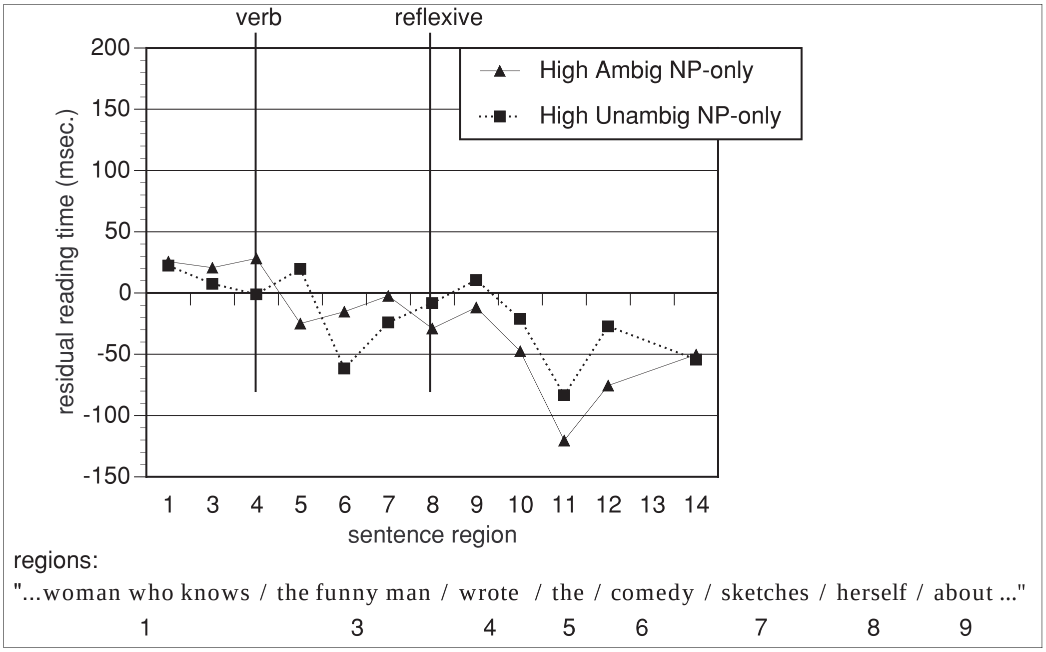


Figure 23: Expt. 2, Low Span Subjects, High NP-only Conditions, n=27

In the high, NP-only conditions, there was no main effects of ambiguity at the word following the reflexive (all p>.1), at the NP in region 3 (all Fs<1) or at the verb (all p>.1). When the two words following the reflexive were combined, there was a main effect of ambiguity ($F_1(1,26)=3.74$, $p<.06$; $F_2(1,19)=4.25$, $p<.05$), with the *un*ambiguous conditions read more slowly than the ambiguous conditions. Thus, there is no evidence for any low attachments in the NP-only condition. This is consistent with the fact that there is no actual ambiguity in the NP-only conditions.

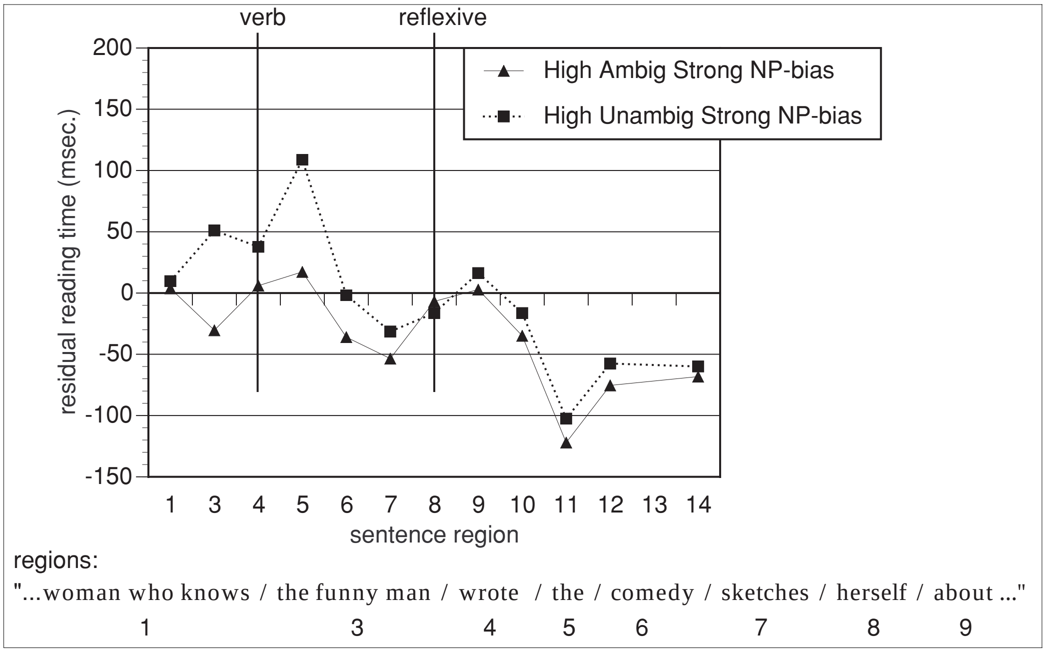


Figure 24: Expt. 2, Low Span Subjects, High Strong NP-bias Conditions, n=27

In the high, strongly NP-biased conditions, there was no main effect of ambiguity at the word following the reflexive (all Fs<1), suggesting that the low span subjects consistently made the high attachments in the strongly NP-biased conditions.

There was a marginal main effect of ambiguity at the NP in region 3 in the item analysis ($F_1(1,26)=2.64$, $p>.1$; $F_2(1,19)=3.75$, $p<.06$), with the unambiguous conditions read more slowly. There was a significant main effect of ambiguity at the verb ($F_1(1,26)=5.64$, $p<.05$; $F_2(1,19)=5.90$, $p<.05$) and at the word following the verb ($F_1(1,26)=3.63$, $p<.06$; $F_2(1,19)=4.62$, $p<.05$), with the *un*ambiguous conditions again read more slowly.

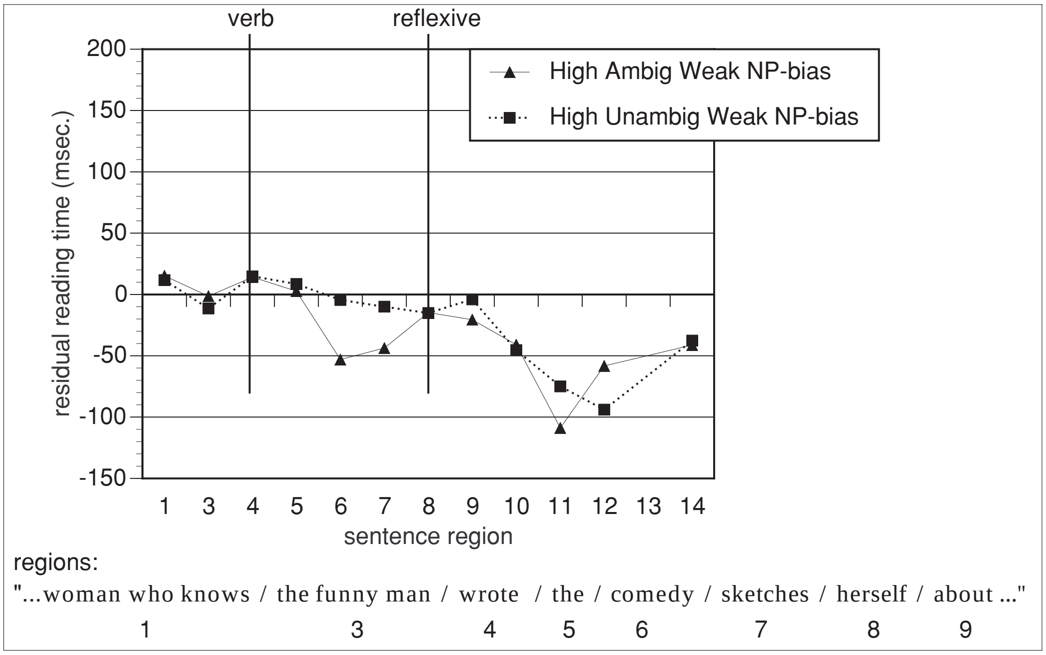


Figure 25: Expt. 2, Low Span Subjects, High Weak NP-bias Conditions, n=27

In the weakly NP-biased high conditions, there was no main effect of ambiguity at the word following the reflexive (F<1). This finding suggests that the low span subjects consistently made high attachments in this condition. There was no main effect of ambiguity at the verb (F<1). There was also no main effect at the NP in region 3 (all Fs<1). Just as in the NP-only conditions and the strongly NP-biased conditions, the effect of ambiguity at regions 6-7 was marginally significant in the subject analysis, but not significant in the item analysis ($F_1(1,26)=3.18$, $p<.08$; $F_2(1,18)=2.25$, $p<.15$). It is not obvious why there should be any effect of ambiguity in these regions.

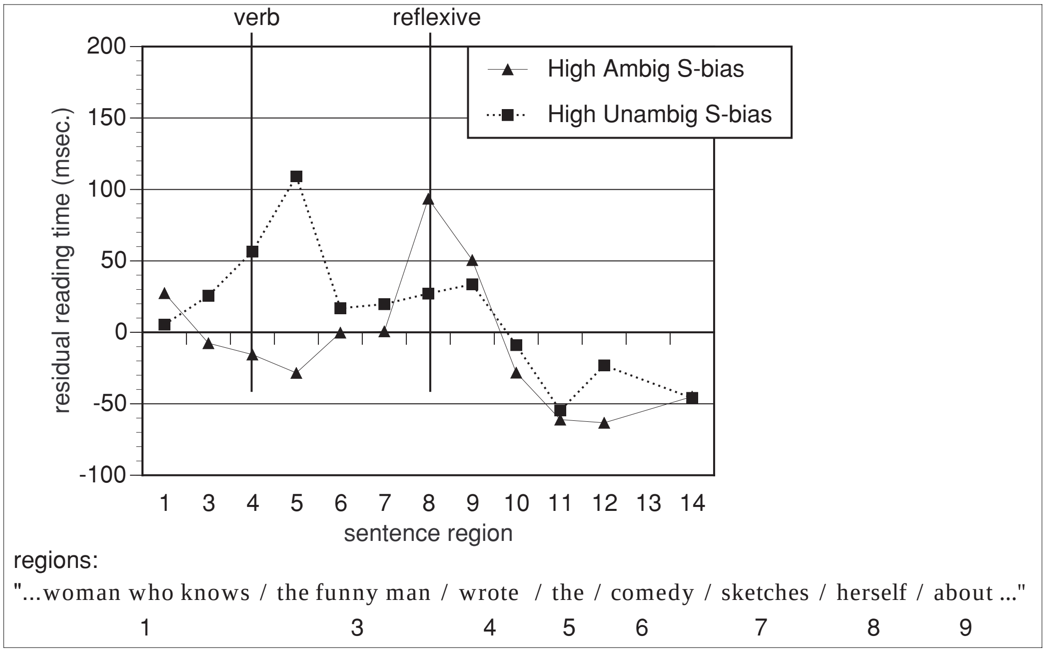


Figure 26: Expt. 2, Low Span Subjects, High S-bias Conditions, n=27

In the high S-biased conditions, there was no main effect of ambiguity at the word following the reflexive, (all Fs<1), although there was a main effect at the reflexive ($F_1(1,26)=3.94, p<.05$; $F_2(1,19)=5.40, p<.05$) and when the reflexive and the following word (R8-9) were combined ($F_1(1,26)=4.17, p<.05$; $F_2(1,19)=4.05, p<.05$). There was no main effect of ambiguity at the verb (all p>.1). However, when the verb and the following word (region 4 and 5) were combined, there was a significant main effect of ambiguity ($F_1(1,26)=12.64, p<.01$; $F_2(1,19)=11.97, p<.01$). There was a marginal main effect of ambiguity at the NP in region 3 ($F_1(1,26)=3.20, p<.08$; $F_2(1,19)=3.07, p<.09$).

5.3.4.2.2 Verb Class Comparison

For the low span subjects, there were no significant interactions between ambiguity and verb class among the low conditions at the word following the reflexive (R9), at the two words following the reflexive (R9-10) (all Fs<1), or at the NP in region 3 (all p>.1). There was a marginally significant interaction at the verb in the item analysis, but no significant interaction in the subject analysis ($F_1(2,26)=2.02, p>.1$; $F_2(2,56)=2.33, p<.10$) When the verb and following word were combined, there was no interaction between ambiguity and verb class (all Fs<1).

In the high conditions, there were no significant ambiguity x verb class interactions at the word following the reflexive, at the NP, or at the verb(all $p>.1$). However, when the verb and the following word were combined, an interaction between ambiguity and verb class was found ($F_1(2,26)=2.79$, $p<.07$; $F_2(2,56)=3.88$, $p<.05$). Thus it appears that verb class is generally less important for the low span subjects than it is for the high span subjects.

In contrast to the high span subjects, the low span subjects showed no significant ambiguity x attachment site x verb class interaction at the word following the reflexive (all $p>.1$). However, there were marginal interactions in the subject analyses at the reflexive ($F_1(3,26)=2.44$, $p<.09$; $F_2(3,75)=2.00$, $p<.14$) and when the reflexive and the following word are combined ($F_1(3,26)=2.98$, $p<.06$; $F_2(3,75)=2.04$, $p<.14$). The interactions were essentially the same when the NP-only conditions are excluded: the interaction was non-significant at the word following the reflexive (region 9) (all $Fs<1$), and was marginally significant in the subject analyses at the reflexive ($F_1(2,26)=2.35$, $p<.10$; $F_2(2,56)=1.93$, $p<.15$) and when the reflexive and the following word were combined ($F_1(2,26)=2.72$, $p<.07$; $F_2(2,56)=1.90$, $p<.15$). As it did for the high span subjects, this interaction indicates the cost of ambiguity differs depending on attachment site and verb class. In other words, when the verb bias and attachment site agree, there is a relatively small ambiguity effect, while when verb bias and attachment site are in conflict, there is a large ambiguity effect. This means that the parser generally follows the analysis that is predicted by the verb bias in the ambiguous conditions and resists revision. In the unambiguous sentence, the parser follows the only analysis allowed by the syntax, even if it is in conflict with the verb biases.

5.3.4.2.3 Comprehension Questions

Table 6 shows the mean scores on the comprehension questions for the low span subjects. As we saw in the ambiguity effects at the disambiguation, verb class appears to have no effect in the low conditions; there was no main effect of verb class among the low conditions (all $Fs<1$), although there was a main effect of ambiguity ($F_1(2,26)=12.66$, p<.01; $F_2(2,56)=12.94$, p<.01). In the high conditions, there was a significant main effect of verb class ($F_1(3,26)=8.29$, p<.001; $F_2(3,75)=5.39$, p<.01), but no main effect of ambiguity (all Fs<1). This suggests that following the verb bias (e.g. high attachment for NP-biased verbs) is easier in the high conditions, while not following the verb bias causes difficulty.

Table 6: Expt. 2 Mean Comprehension Question Scores for Low Span Readers

	Ambiguous	**Unambiguous**
Low Strong NP-bias	72%	86%
Low Weak NP-bias	74%	80%
Low S-bias	73%	84%
High NP-only	90%	87%
High Strong NP-bias	84%	86%
High Weak NP-bias	78%	80%
High S-bias	71%	76%

When both attachment sites were included in the analysis, there was an interaction between verb class and attachment site ($F_1(3,26)=2.82$, p<.06; $F_2(3,75)=3.25$, p<.05). This interaction was also present when the NP-only verbs were removed from the analysis ($F_1(2,26)=2.58$, p<.08; $F_2(2,56)=3.06$, p<.05). This confirms that the verb bias affects the different attachments differently, with attachments that follow the verb bias generally being easier than those that go against the bias.

5.3.4.3 Subject Group Comparison

When the high and low subjects were compared directly, significant differences were found at the word following the reflexive (region 9). There was no overall main effect of subject group at the word following the reflexive (F<1). However, there was a significant interaction between subject group and attachment site ($F_1(1,54)=4.03$, p<.05; $F_2(1,58)=3.01$, p<.09). There was also a marginally significant 3-way interaction between attachment site, subject group, and ambiguity ($F_1(1,54)=2.84$, p<.10; $F_2(1,58)=3.01$, p<.09). There were no other significant interactions involving subject group and attachment site (all p>.1)

When the high attachment conditions were analyzed separately, there was a marginally significant main effect of subject group ($F_1(1,54)=3.00$, $p<.09$; $F_2(1,58)=3.44$, $p<.07$). In the high conditions, there were no interactions with subject group, nor were there any main effects or interactions within individual verb classes (all $p>.1$).

In the low conditions, there was no main effect of subject group ($F<1$), but there was a significant interaction between ambiguity and subject group ($F_1(1,54)=4.47$, $p<.05$; $F_2(1,58)=4.73$, $p<.05$). In the low S-bias conditions, there was a significant ambiguity x subject group interaction ($F_1(1,54)=4.64$, $p<.05$; $F_2(1,19)=4.95$, $p<.05$). There were no significant interactions between ambiguity and subject group in either the strongly or weakly NP-biased conditions (all $p>.1$). A main effect of subject group was only found in the low ambiguous S-bias condition ($F_1(1,54)=3.22$, $p<.08$; $F_2(1,19)=4.61$, $p<.05$); there was no effect of subject group in the low unambiguous S-bias conditions (all $p>.1$).

There was no main effect of subject group at either region 3 or 4 (all $F<1$), nor were there any interactions involving subject group at either of these regions (all $F<1$).

In summary, the analysis of subject groups shows a significant difference between the high span and low span subjects, which appears to be due to the low S-bias conditions, where the high span subjects read the word following the reflexive more quickly than the low span subjects.

5.3.5 Summary of Experiment 2 Findings

Table 7 shows a summary of the effect of ambiguity at the disambiguating regions in Experiment 2. As can be seen, there is a significant 3-way interaction between ambiguity, attachment site, and verb class for the high span subjects (see 5.3.4.1 /Verb Class Comparison for details) and a marginally significant interaction for the low span subjects (see 5.3.4.2./Verb Class Comparison for details) The high span subjects consistently make high attachments in the NP-biased conditions, much as they did in Experiment 1. In the S-biased conditions, the high span subjects appear to be consistently making the low attachment.

Table 7: Cost of Ambiguity at Disambiguation in Experiment 2.
All effects measured at the word after the reflexive (region 9) unless otherwise specified.
*=significant p<.05; #=marginally significant p<.10

	High Span Subjects		**Low Span Subjects**	
	Low Cond.	High Cond.	Low Cond.	High Cond.
NP-only	n/a	16	n/a	-21
Strong NP-bias	65*	22	110*	-14
Weak NP-bias	57* (R9-10)	-43*	55* (R8-9)	-17
S-bias	13	92*	66#	66* (r8)

In the NP-biased conditions, the low span subjects behaved similarly to the high-span subjects, where they made the high attachment in the ambiguous conditions. This is different from Experiment 1, where the low span subjects showed evidence for both high and low attachments in NP-biased conditions. While Experiment 1 showed ambiguity effects in the high and low conditions, the cost of ambiguity was much greater in the low conditions than in the high conditions, just as in Experiment 2. In both experiments there was a large cost of ambiguity in the high NP-biased conditions, which indicates a strong tendency for high attachments. Where the experiments differ is in the presence or absence of a relatively small cost of ambiguity in the high conditions.

In the S-biased conditions, the low span subjects showed ambiguity effects in both the high and low conditions, which suggests that they made both high and low attachments in the ambiguous conditions. However, despite this variability, the results suggest a tendency to respect the bias of the embedding verb, since the ambiguity effect in the low S-bias conditions was only marginally significant, while it was significant in the high conditions.

5.3.5.1.1 Pronoun Costs

As noted earlier, in many conditions the pronoun in region 3 of the unambiguous stimuli was read more slowly than the full NP in the ambiguous stimuli. A number of possible explanations were mentioned for this effect. An examination of the data from all of the conditions allows the question of why the pronoun in region 3 is frequently read more slowly than the full NP to be properly addressed. In Figure 27, it can be seen clearly that the high span subjects had trouble with the unambiguous pronoun precisely when it provided information that was contrary to the information from the embedding verb. In particular, the pronoun cost was highest in the low strongly NP-biased condition and in the high S-biased conditions, and it was weakest (i.e. the pronoun was read faster than the full NP) when the pronoun was consistent with the verb bias information (i.e. the high NP-only condition and the low S-biased condition). The pronoun cost is shown for region 4 in the high S-bias conditions (the dashed lines) because the cost is much higher at the word following the pronoun in these conditions. In all other conditions, the cost of the pronoun was lower at the region following the pronoun (the verb), and in many conditions the effect actually reversed itself at that region, with the ambiguous condition (the full NP) read more slowly than the unambiguous condition (the pronoun). These effects suggest that the effect of the pronoun is not fully felt in the high S-bias conditions until at least the word following the pronoun, while in the other conditions it is seen immediately at region 3.

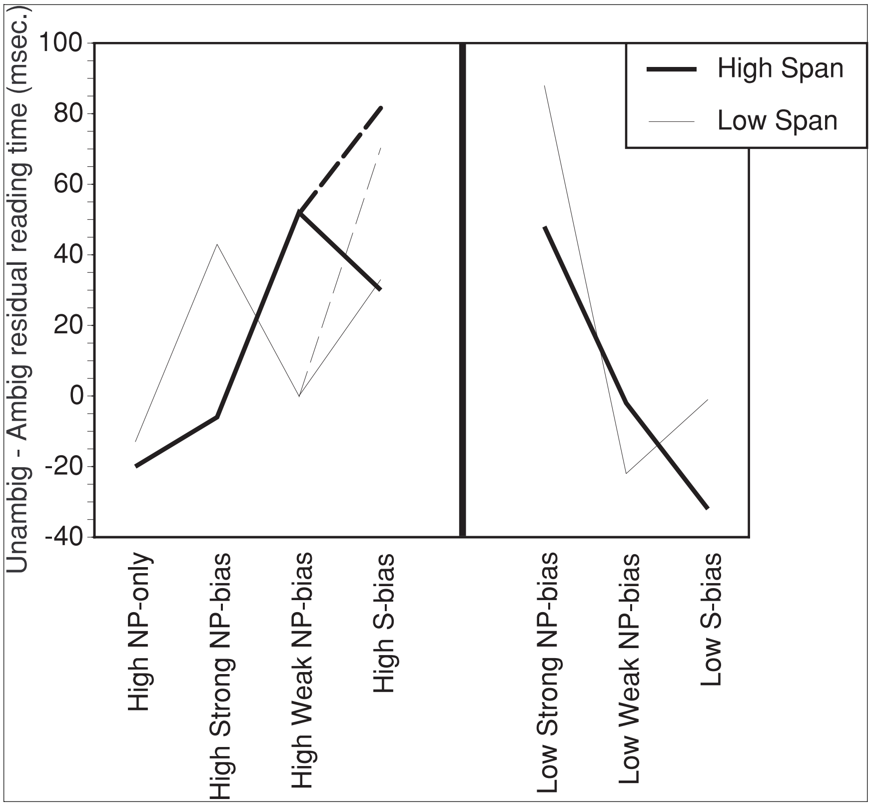


Figure 27: Expt. 2, Cost of Pronoun vs. Full NP at Region 3 (Unambig. - Ambig.)
The dashed line is the effect at region 4.

The findings for the low span subjects are similar to those of the high span subjects, except that the low span subjects seem to not slow down for the pronoun in either of the weakly NP-biased conditions. It is not immediately obvious why the low span subjects should behave differently in the weakly-biased conditions, unless the fact that the verb bias is so weak means that they form no expectations whatsoever for new material, so nothing surprises them.

The variation across verb classes found here is quite comparable to that seen in Trueswell et al. (1993), though it manifests itself in different ways in the two experiments (sentences in Trueswell, et al. (1993) were disambiguated differently). The common effect is that continuations inconsistent with the verb bias were slowed significantly immediately following the verb.

5.4 Discussion

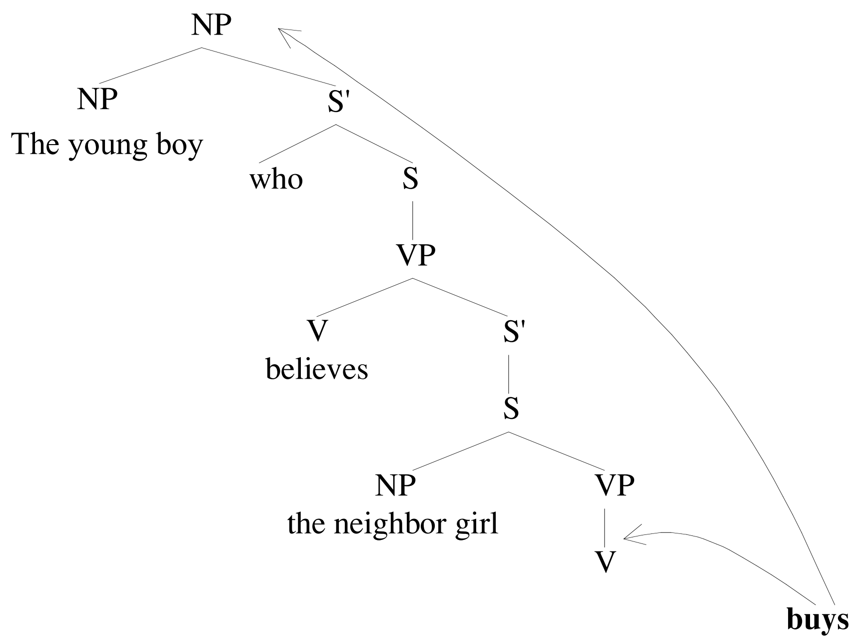
The high span subjects performed consistently in both experiments: in the NP-biased conditions (both weak and strong NP bias), they showed strong evidence for high attachments in the ambiguous conditions, while in the S-biased conditions they showed strong evidence for low attachments.

Table 8: Cost of Ambiguity for High-span Subjects
All effects at region 9 unless otherwise noted, *= significant p<.05

	Expt. 1		**Expt. 2**	
	Low Cond.	High Cond.	Low Cond.	High Cond.
NP-only	n/a	n/a	n/a	16
Strong NP-bias	107*	2	65*	22
Weak NP-bias	97*	0	53* (R9-10)	-43*
S-bias	n/a	n/a	13	92*

A very simple explanation can be given for this behavior: the subjects obey the Reanalyze as a Last Resort strategy. In the NP-biased conditions, the initial attachment of the NP (region 3) respects the bias, and the following verb (region 4) is attached as the verb of the higher clause, with no need for reanalysis. Effects of RALR can also be seen in the S-biased conditions if we assume that the ambiguous NP in the S-biased conditions is initially analyzed as the subject of a sentential complement. If the NP is already the subject of an embedded clause, no reanalysis is required for low attachment. In fact, if the NP is initially analyzed as the subject of an embedded clause, the high attachment should not even be considered, low attachment is the only option that does not require reanalysis. The example in (228) shows that if the verb were analyzed high, the NP *the neighbor girl* would have to be reanalyzed from a sentential subject to a simple direct object.

(228)



Thus, the high attachments in the NP-biased conditions and the low attachments in the S-biased condition can be accounted for by the combination of Reanalysis as a Last Resort and the immediate use of verb frequency information.

The behavior of the low span subjects in the two experiments cannot be explained so straightforwardly. The ambiguity effects for the low span subjects in both experiments is summarized in Table 9 below. The boxed areas are those that do not follow the predictions of the model suggested above for the high span subjects (i.e. RALR and follow verb-bias). The major puzzle shown by the table is why the low span subjects in Experiment 1 show evidence for low attachments in the NP-biased condition, while the low span subjects in Experiment 2 show no such evidence.

Table 9: Cost of Ambiguity for Low Span Subjects
All effects at region 9 unless otherwise noted
*=significant $p<.05$; #=marginally significant $p<.10$

	Expt. 1		**Expt. 2**	
	Low Cond.	High Cond.	Low Cond.	High Cond.
NP-only	n/a	n/a	n/a	-21
Strong NP-bias	168*	49*	110*	-14
Weak NP-bias	102*	63*	55* (R8-9)	-17
S-bias	n/a	n/a	66#	66* (R8)

Given that there is strong evidence from the high span subjects for Reanalysis as a Last Resort, the explanations of the differences for the low span subjects will assume that RALR is functioning in the parser for all subjects. I assume that the difference between the two subject groups is not related to RALR, but is instead related to what information is available to the parser and what other constraints it is under (see 5.4.1.1 below for evidence showing reduced availability of information to low span subjects and for a claim that they have fewer computational resources). Two different possible explanations for these anomalous findings will be presented here, which both involve an effect of memory limitations on which structures are available to the parser, thus making the effects of RALR differ from those seen in the high span subjects. Common to both explanations is the assumption that the low span subjects make less use (but not no use) of frequency information. Possible reasons for why the low span subjects make less use of frequency information include: they don t store frequency information as efficiently, they have more difficulty accessing frequency information, or they have more difficulty using frequency information to guide parsing decisions. There is no evidence from these experiments about which of these options is correct, so I will take no stand on the issue of why these subjects might make less use of frequency information. The relationship of this work to other work on memory span and parsing ambiguity is discussed below in 5.4.1.1 .

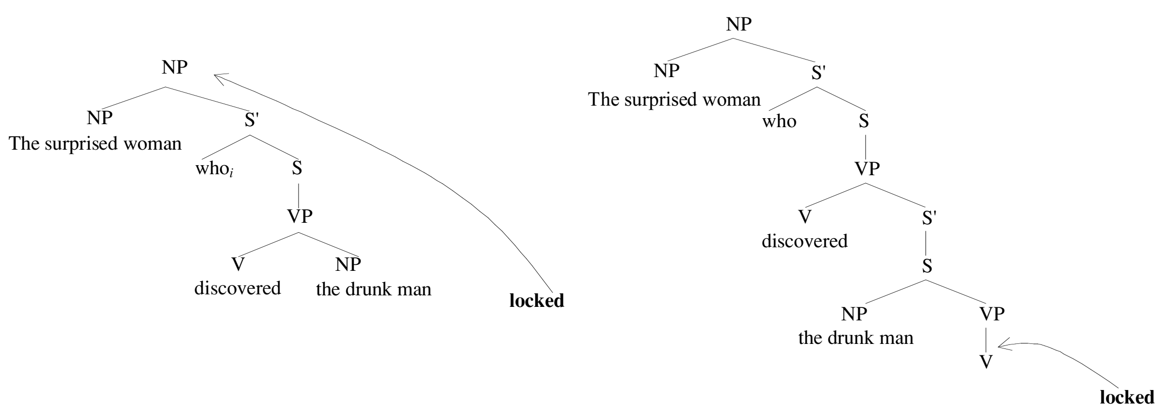
If verb bias information is not used as well by the low span subjects to guide parsing decisions, they should perform with more variability (within a given verb class) than the subjects who are able to make full use of the information. I interpret the pronoun costs in Figure 27 as a surprise effect when the unambiguous pronoun is in conflict with the parser s expectations. Extending this line of reasoning, I assume that the ambiguous full NPs are attached in accordance with the verb bias, since they show no effects of surprise. From this interpretation of Figure 27 I conclude that the initial attachment of the NP/pronoun (region 3) is guided by verb bias (e.g. if the verb is S-biased, then the NP will be attached as subject of a sentential complement). If there is no biasing information available to the parser, the two subcategorizations should appear equally strong, and as a result both NP-complement and S-complement analyses should be available. In each of the explanations presented below, this relative lack of verb-bias utilization helps explain why there is a cost of ambiguity in both the high and the low S-bias conditions, and further predicts that there should be ambiguity effects in all of the NP-biased conditions as well.

Thus, if verb bias information is generally less useful for the low span subjects, these subjects should show a cost of ambiguity in all conditions with ambiguous embedding verbs, and whether the region 3 NP is a full NP or a pronoun should have no effect at the reflexive in the NP-only conditions, since they are not syntactically ambiguous. The only facts that this does not account for is the lack of ambiguity effects in the high NP-biased conditions of Experiment 2. The two explanations that follow are attempts to explain why in Experiment 2 there is evidence in the NP-biased conditions for only high attachments, but in Experiment 1 there is evidence for both high and low attachments.

In both accounts of the differences between the two experiments, the crucial factor is the fact that the ambiguities in Experiment 2 were embedded one clause deeper than the ambiguities in Experiment 1. The first account for the differences between the two experiments is that deeper embedding causes more processing difficulty than shallower embedding, and this causes the parser to prefer shallower embeddings when approaching its processing limit. For the low span subjects, this limit is lower than for the high span subjects. The sentences in Experiment 2 do not approach the threshold of the high span subjects, while they do approach the threshold for the low span subjects, so the low span subjects should show a depth-of-embedding effect. The consequence of this extra depth for the low span subjects is that when they are processing more deeply embedded structures, there is a preference to make attachments that will leave the parser in a less deeply embedded state. This effect could be termed buoyancy. Under this conception, buoyancy comes into play when the processor is under significant load, as in the deeply-embedded items in Experiment 2. The effect of buoyancy is that if there are competing analyses, one of which involves attachments higher in the tree (effectively closing off the lower structure) than the other, the analysis involving the higher attachments will be pursued. The effect of buoyancy is illustrated in the examples in (229) and (230).

(229) the surprised woman who discovered the drunk man locked

(230)



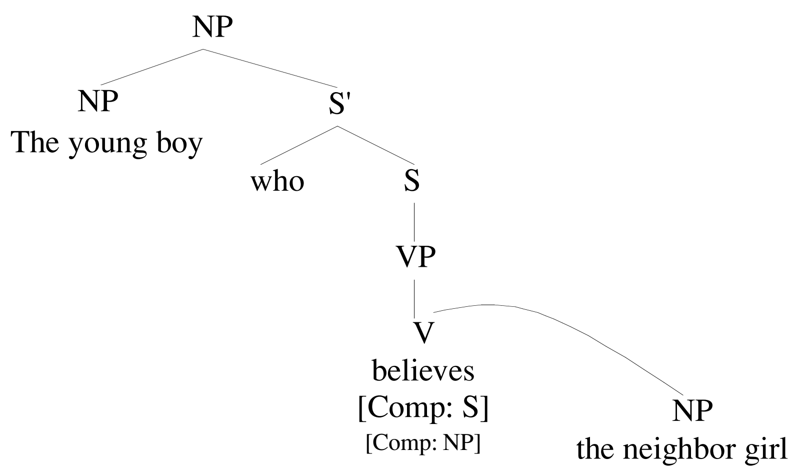
If both of the analyses in (230) are under consideration in parsing the verb *locked* in (229), the high attachment in the left tree should be pursued because it leaves the parser in a less embedded state. A general prediction of this model is that putting the parser under a heavy load should result in more Early Closure Effects.

The second account of the lack of low attachments in Experiment 2 is based on a different problem that might occur when low span subjects are parsing more deeply embedded structures. The idea is that when the parser is under stress, it is less able to posit new syntactic nodes. In effect, Minimal Attachment becomes more important when the parser is under significant stress. Under this theory, in Experiment 2, where the ambiguity is embedded one level deeper in the tree, the computational system has fewer resources available, and is more likely to make minimal attachments (i.e. NP complements/high attachments).

If Minimal Attachment is the only strategy being used, the prediction is that there should be relatively little evidence for low attachments in any conditions (not even in the S-bias conditions, since attaching the NP as a sentential subject requires more structure than attaching as a simple NP-complement). However, recall the prediction that the restricted use of verb bias information available to the low span subjects in the NP-biased conditions should allow at least a few low attachments (S-complements) and there should be at least a few NP-complements (high attachments) in the S-biased conditions. When Minimal Attachment and weak verb bias utilization are combined, the tension between the two appears to yield the correct results. Consider what would happen in Minimal Attachment raises the number of high attachments by 10% and weakened verb bias allows 10 % more of the dispreferred attachments. In the NP-biased conditions, the weakened bias information should raise the number of S-complements by 10%, while Minimal Attachment lowers the number of S-complements by 10%. The net result is that these two effects cancel each other out, and the initial verb bias information is respected (i.e. NP complements/high attachments are pursued). In the case of the S-biased verbs, Minimal Attachment again predicts 10% more NP-complements, and weakening of the S-bias should also cause 10% more NP-complement parses. The result in this case is that there should be 20% more NP-complements than the verb bias predicts. This effect matches that seen in Experiment 2, where the S-biased conditions showed evidence for high attachments and low attachments, while the NP-biased conditions only showed evidence for high attachments (i.e. they followed the verb bias).

This second explanation should be fairly easy to implement within the computational model proposed in this dissertation. When the parser is under increased stress (e.g. in deep embeddings), it becomes more difficult to posit new nodes. This could be straightforwardly implemented by requiring extra attempts at finding a simple attachment before any new nodes (predicted heads or other null heads) are posited. For example, if *the young boy who believes* has already been parsed and *the neighbor girl* is received (as in (231) below), there will initially be no way to attach *the neighbor girl* because the verb prefers a sentential complement (recall that *believe* is S-biased). Due to the strong preference for the sentential complement, the parser will initially have no access to the information stating that *believe* can also take an NP-complement. The normal action for the parser would be to attach extra material to the NP so that it could fulfill the role of a sentential complement. However, since there is now an extra cost for building new nodes, the verb will be checked again (for its full range of possibilities, not just the preferred one), this time yielding the fact that it can also take an NP-complement. Once this has been seen, the NP can be attached as an NP-complement without requiring any new nodes.

(231)



5.4.1.1 Other Reading Span Results

Reading memory span has been implicated in ambiguity resolution in a number of papers by Just, Carpenter, and colleagues (Just and Carpenter 1992, MacDonald, Just and Carpenter 1992 and Pearlmutter and MacDonald 1995). The two papers by MacDonald and colleagues involved the resolution of main verb/reduced relative ambiguities, such as those shown in (232) and (233), in which *warned* can be either a matrix verb (232) or the verb of a reduced relative clause (233).

(232) The experienced soldiers warned about the dangers before the midnight raid.

(233) The experienced soldiers warned about the dangers conducted the midnight raid.

The two papers by MacDonald and colleagues are very similar in their analysis of the phenomena discussed here, and I will accordingly treat them as one analysis. In the analysis of their results, they assume that in the absence of contextual/pragmatic biases, simple verb frequency favors the active (main) verb analysis. In the matrix verb condition ((232)), they found that high span subjects consistently showed a slowdown at the point of disambiguation (raid.) in the ambiguous conditions, while low span subjects showed no such ambiguity effect. In the comprehension questions they found no difference in accuracy between the high and low span readers, but they did find that the ambiguous conditions had a higher error rate than the unambiguous conditions. In the dispreferred alternative (reduce relative) condition, they found that the high span subjects showed a large slowdown due to ambiguity, while the low span subjects showed a smaller, though still significant effect. As in the matrix condition, they found that the unambiguous conditions were answered more accurately, and they further found that high span subjects were better able to answer the comprehension questions than the low span subjects (who were near chance in the ambiguous conditions).

In the context of their parallel parsing model, they interpreted these results to mean that the high span subjects were better able to keep multiple parses in memory, while the low span subjects were more likely to discard the less preferred parse. The larger ambiguity effect for the high span readers is a result of the fact that these subjects have more parses to sift through in deciding which one to keep upon disambiguation. In the preferred continuation, the high span subjects needed to decide that the other interpretation was not worth pursuing any further while the low span subjects were able to continue with the single parse that they were pursuing. In the dispreferred continuation (the reduced relative), the high span subjects took a long time to decide that their dispreferred parse should be promoted to the most preferred parse, while the low span subjects saw that the one parse they had was not consistent with the sentence and immediately engaged in repair/backtracking. The effect of this was that the low span subjects showed a significant cost of ambiguity in the dispreferred conditions, but a smaller one than the high span subjects. In short, they found that the high span subjects showed a large effect of ambiguity in both the preferred and dispreferred continuations, while the low span subjects only showed ambiguity effects in the dispreferred condition. This follows the predictions of their model in which high span subjects should be more able to keep multiple parses in memory than the low span subjects, who are predicted to behave more or less as if they had a serial parsing system.

It is interesting to note that MacDonald, Just, and Carpenter (1992) did not find any ambiguity effects for the high span subjects during the ambiguous region in the matrix verb condition. This is surprising since they did find an ambiguity effect at the point of disambiguation. This difference between the ambiguous and unambiguous regions is not predicted by their model. While the general results are largely consistent with the explanation offered in MacDonald, Just, and Carpenter (1992) and Just and Carpenter (1992), they are also largely consistent with the parsing system discussed in this thesis. Although the SPARSE model predicts that only the main verb analysis would be constructed, the pattern of ambiguity effects following a region without ambiguity effects is exactly what would be predicted by this model if the reduced relative analysis were pursued. If (for whatever reason) high span subjects are able to construct the reduced relative, they should show no slowdown in the ambiguous regions and should show an ambiguity effect at the disambiguation, when they are forced to adopt the main verb interpretation of the verb. The low span subjects, on the other hand, might have more difficulty positing the extra nodes needed for the reduced relative interpretation of the verb, and would thus consistently pursue the main verb attachment. This would result in no effect of ambiguity at the disambiguation, just as was found in their results.

The results of Pearlmutter and MacDonald (1995) also show that high span subjects are more sensitive to extra-grammatical constraints than low span subjects. After checking the plausibility of the matrix and reduced relative interpretations of their stimuli, Pearlmutter and MacDonald (1995) found that their high span subjects were more able to take advantage of plausibility differences than the low span subjects. In a similar task that manipulated animacy, Just and Carpenter (1992) also showed that high span subjects are able to rapidly take advantage of plausibility information, specifically animacy cues. These findings of rapid use of plausibility information are relevant since this sort of extra constraint is similar to the verb biases that were manipulated in these experiments, in that both are knowledge not strictly needed for the computation of grammaticality that is used more fruitfully by high span subjects than low span subjects.

In sum, the results in MacDonald, Just and Carpenter (1992), Pearlmutter and MacDonald (1995), and Just and Carpenter (1992) show that high span subjects are better able to make use of extra-grammatical knowledge than low span subjects. The data from MacDonald, Just and Carpenter (1992) and Pearlmutter and MacDonald (1995) also seem to indicate that high span subjects are able to consider a wider range of syntactic structures than low span subjects. Both of these findings are consistent with the parsing theory presented in this thesis.

5.4.1.2 Implications for Monotonicity Accounts

The findings of these experiments are also relevant to the question of how syntactic structure is represented and computed. This section discusses a number of parsing effects including the results from these experiments, and then discusses the consequences of this constellation of facts for D-theory (Marcus et al. 1983) and successive accounts of monotonic structure building (Weinberg 1993, Gorrell 1995, but see also Sturt and Crocker 1999 for a different formulation of a monotonicity account).

Two findings from the experiments are important for this discussion: there is a general tendency to not reanalyze, and there is a cost associated with reanalyzing the structures used in these experiments (at region 8). In particular, there is a preference (in the NP-biased conditions) to analyze the ambiguous verb (phrase) as a main verb and a corresponding cost to lower it from a main verb analysis to a sentential complement interpretation. Likewise, there is a preference (in the S-biased conditions) to analyze the ambiguous verb as part of a sentential complement, which is shown through a significant cost to raise the V(P) from a sentential complement position to a higher position.

Consider also the following garden path sentence taken from Pritchett (1992).

(234) The woman gave the man who was racing the car.

The problem with this sentence is that people initially interpret the NP *the car* as the direct object of *racing*, and have difficulty analyzing it as the theme of *gave*. The fact that people have this difficulty shows both that people have a preference to attach to a more recent site (e.g. *racing*) than less recent sites (e.g. *gave*), and it also shows that the change required to change an attachment (e.g. move *the car* from *racing* to *gave*) also has a significant cost. Similar recency effects can also be shown for adjuncts, as exemplified by the time adjunct in the following sentence

(235) I heard that Dorothy was caught in a tornado last week.

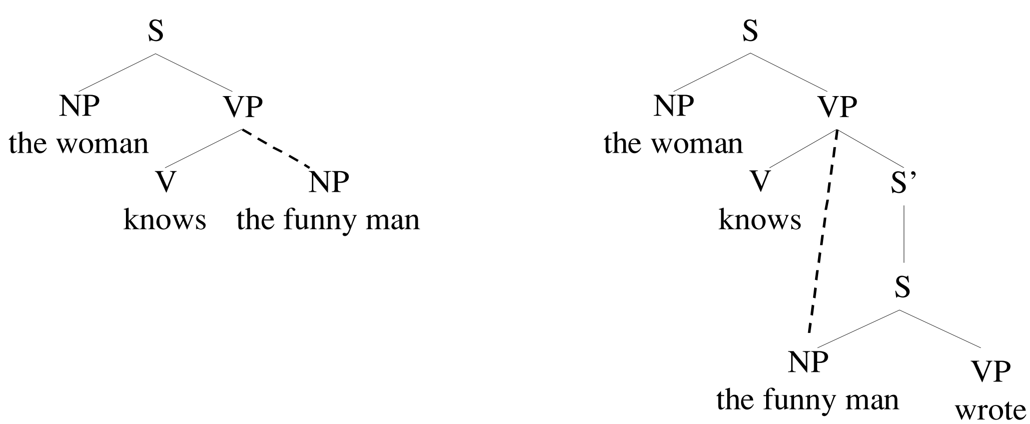
The preferred interpretation for *last week* is as a modifier of *caught* rather than as a modifier of *heard*, and it is clearly more difficult to interpret *last week* as a modifier of *heard*. Phillips and Gibson (1997) showed evidence that a local adjunct attachment is preferred to a less local argument position, as seen in the following example:

(236) Although Erica hated the house she had owned (it) for ten years

When the object pronoun *it* is used to disambiguate towards a matrix subject analysis of *she owned ,* Phillips and Gibson showed that there is a significant slowdown at the disambiguating regions relative to an unambiguous control. In contrast, when *it* is omitted and *she* is analyzed as part of a relative clause modifying *the house*, there is no slowdown compared to an unambiguous control. These results again show that there is a preference to attach incoming words to recently processed material. These last few examples show that there is a general preference for locality, and furthermore that there is a cost for reanalyzing away from the preferred structure (even in the cases where the most local attachment is not the preferred one).

With these effects in mind, I now turn to D-theory and show how it accounts for reanalysis costs, and how the facts presented above pose a challenge for an interesting proposal of D-theory. As discussed in /1.2.3.2 , D-theory and its successors provide a very elegant account of why the standard NP-S ambiguity is so easy (the ease of reanalysis was shown experimentally in Sturt, et al. 1999b). In D-theory, the changes needed to switch an NP from a direct object reading to a sentential complement reading do not actually constitute reanalysis at all, but are instead part of the regular monotonic structure building operations. Domination (not *immediate* domination) is the structural primitive manipulated by D-theory parsers, and adding domination statements to a tree description is part of the automatic operations performed by the D-theory parser. Because it is done by the automatic parser, adding domination statements is thought in D-theory to be completely cost-free. Raising elements in a tree, on the other hand, is very costly, because it requires removing domination links from the existing representation, and removal of information is not one of the automatic operations of the parser.

(237)



The trees in (237) show the structural change involved in changing a NP from a direct object analysis to an S-complement analysis. As can be seen by the dotted line, the domination statement between the VP and the NP *the funny man* does not need to be retracted in order to lower the NP into the embedded subject position. For this reason, the standard NP-S temporary ambiguity is correctly predicted to be very easy, compared to examples in which a domination statement needs to be retracted, as in the following example from Marcus (1980).

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

In this example *the sock* is initially taken to be an object of *mending*, and the change necessary to make it the matrix subject requires elimination of the domination statement from the VP headed by *mending* to the NP *the sock*. It is curious in D-theory that even though the change necessary to reanalyze in the sentences in these experiments is very easy (as shown in (237)), that option is not taken, and instead a less local option is taken, thereby violating the recency preference that appears to be prevalent in the parser.

Parsing preferences are accounted for in D-theory by assuming that the preferred reading of a sentence will correspond to the standard referent, which is formed by making all of the domination links in the tree description maximally short. This means that all attachments will be interpreted at the highest location permitted by the tree description. To account for the recency preferences noted above, a D-theoretic parser would have to posit domination links from the lower attachment sites. However, a simple strategy of always asserting dominance from the lowest attachment site will not account for the data from the both the strongly and weakly NP-biased conditions of these experiments. Recall that the NP-biased conditions showed a preference for the high attachment site that cannot be accounted for by a consistent preference to attach as low as possible.

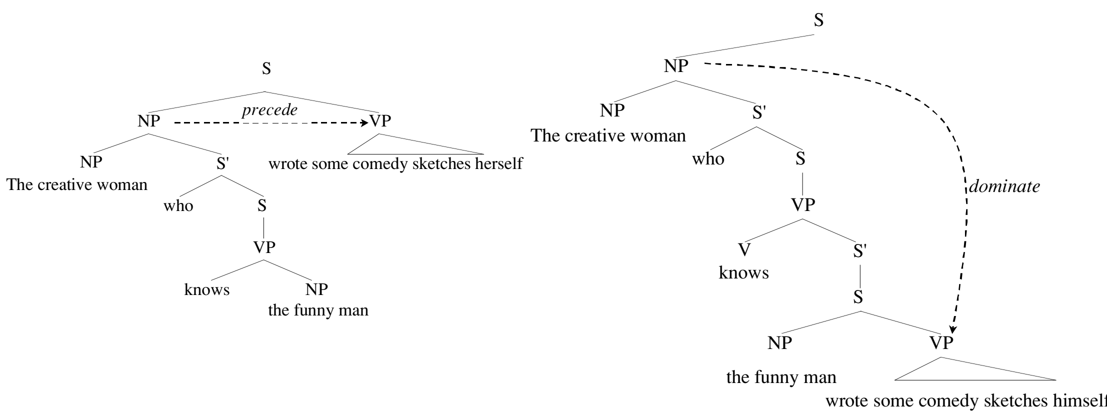
The preferences found in these experiments can be accounted for by a theory in which nodes are always attached as high as possible (e.g. Minimal Commitment Theory as discussed in Weinberg 1993). If the ambiguous verb is attached as high as possible, the observed preference for high attachment in these experiments is accounted for. However, the general preference for low (recent) attachment sites (as in (234) and (235)) is not accounted for. An obvious solution to the problem from our findings that locality is avoided in these experimental conditions is to assume that there is cost for reanalysis. However, this option is at odds with the general position in D-theory that lowering (is needed for the standard NP-S ambiguity) requires no reanalysis. Without adopting the assumption that there is some cost for analysis in these examples, it is difficult to reconcile these findings with the claims of D-theory.

However, even if a solution can be found for the parsing preferences, there is another problem for D-theory in these experiments. Recall that in the NP-biased conditions there is a significant cost for reanalysis to the lower position, which is reflected in both reading times and comprehension accuracy. Under a D-theoretic analysis, the only difference between the high and low attachments is that the low attachment contains the extra domination statements necessary to ensure that the verb is interpreted as part of a sentential complement of the embedding verb. The finding from these experiments that lowering the S-complement has a significant cost is not consistent with the D-theory position that the addition of domination statements is a cost-free part of the automatic operations performed by the parser.

Two different approaches might be taken to account for the difficulty of reanalysis found in these experiments. The first tack involves changing the theory so that reanalysis is required in the structures used in these experiments but is not required in the standard NP-S ambiguity, and the second involves redefining what can be done automatically by the parser. In the theory presented in Gorrell (1995), the syntactic primitive of precedence is added to the list of primitives that cannot be retracted by automatic parsing operations. Gorrell, following Partee, et al. (1993), assumes that there must be either a dominance or a precedence relation between every pair of nodes in a syntactic tree. These two parts of Gorrell s theory predict that the only structural change that does not result in a garden path is the addition of new nodes between existing nodes. Any other rearrangement of nodes will require retraction of existing dominance or precedence statements. This constraint on adding new structure successfully accounts for the fact that the verb can easily be attached as part of a sentential complement when it is encountered (even if it forces reanalysis of the NP), but can t be lowered into that position if it has been attached elsewhere previously.

As can be seen in the tree on the left in (239), there is a precedence relation between the highest NP and the VP when the ambiguous verb is given the high analysis. However, when the ambiguous verb is given the low analysis, there is a domination relation between the highest NP and the VP. Because of the fact that the precedence relation must be removed to switch from the high attachment to the low attachment, Gorrell s theory correctly predicts that the reanalysis required in these experiments causes a garden path.

(239)



While Gorrell s theory successfully accounts for the fact that reanalyzing in these sentences is a conscious garden path, it has no account for the initial parsing preferences. Gorrell is careful to state the his parser builds structure according the principles of the grammar, but it does not attempt maximal satisfaction of grammatical principles (Gorrell (1995), p. 100). If Gorrell s parser attempted maximal satisfaction of grammatical principles, as Pritchett s (1992) parser does, then high attachment of the verb would be predicted. High attachment would be preferred because that would provide a theta-role for the highest NP (the head of the relative clause) and would likewise satisfy the embedding verb, whose selectional features can be satisfied by an NP direct object. The constraint Gorrell poses on syntactic structure building is Simplicity: No vacuous structure building (p. 100). This constraint provides no guidance to the parser in this situation, since there is no vacuous structure. Note that under Gorrell s interpretation of vacuous structure, structure is vacuous only if it serves no useful purpose or is not required in the phrase marker. This means that the S /CP node required in the sentential complement is not vacuous, despite the fact that it has no overt content. Thus, neither the high nor the low attachment of the verb contains any vacuous structure, and Simplicity provides no guidance for how to parse the verb.

Even though Gorrell doesn t discuss general recency preferences, there is room within his theory for a recency preference, which would account for the parsing preferences seen in (234)-(236). However, such a recency preference would predict just the opposite of the initial attachment preferences that are seen in these experiments, where the verb is consistently attached high, rather than to the more recent NP. Thus, while Gorrell does predict the difficulty seen in the reanalysis from the high to the low attachment of the verb, he has no account for the fact that the verb is attached high in the first place.

The second possible approach in D-theory to accounting for the difficulty of the lowering reanalysis is to say that the garden path in the sentences in these experiments is caused by something other than retraction of domination statements. The only explicit claim that D-theory makes about garden paths is that the removal of a domination statement should cause a garden path. There is no claim that the *only* possible cause for a garden path is the removal of a domination statement. Although D-theory claims that the addition of domination statements is governed by the automatic operations of the parser and should thus be cost-free, it is not inconceivable that there might be some occasions when the computation required to determine which domination statements to add could be difficult enough to cause garden path effects. Assuming that some account could be found for the initial parsing preferences, a D-theoretic account might rely upon the high cost to determine that the low NP-biased conditions require lowering. Under this account, the difficulty seen in both the low NP-biased conditions and the high S-biased conditions is due entirely to the difficulty in determining what needs to be done. In other words, the problem is simply too difficult for the automatic parser to deal with successfully. This is not to say that all lowering would be difficult. Indeed, under such an account it could still be quite easy for the direct object/embedded subject to be lowered. Instead, the difficulty in the experimental sentences would be in determining if/how to lower the ambiguous *verb,* not in lowering the NP. For this account to be tenable, the claim that the addition of domination links is always easy would need to be weakened, and it would also require an account of the fact that recency affects most attachment decisions, but not those in these experiments.

Summing up for the monotonicity accounts, there appear to be solutions that could explain the fact that the reanalysis required in these experiments is difficult. However, the problem of accounting for the initial parsing preferences appears to be much more difficult, since the NP is not considered to be reanalyzed within D-theory. By contrast, any theory in which reanalysis is assumed to take place in the NP-S ambiguity can straightforwardly account for the fact that the high attachment of the verb is preferred over the low attachment, by invoking some equivalent of the Reanalysis as a Last Resort constraint.

5.5 Conclusions and Implications

The experiments discussed in this chapter were performed to answer a simple question about parsing: is reanalysis of existing structural commitments considered only when all other options have been exhausted, or is it considered before some analyses that do not require reanalysis? Based on the experimental results, the answer to this question seems fairly clear: reanalysis is only considered as a last resort, when all other options have been exhausted. The fact that the sentential complement analysis is apparently not considered in the NP-biased conditions provides strong support for this position. The fact that the sentential complement analysis is consistently pursued in the S-bias conditions is also consistent with RALR, on the assumption that in the S-biased conditions, the ambiguous NP is initially analyzed as an embedded subject.

However, these experiments also raise a number of additional questions. The remainder of this section outlines some of these questions and speculates about possible answers to these questions. One question raised by the experiments is how lexical biases influence parsing. In the parsing algorithm presented in Chapter 2 and Chapter 3, structure is built with the knowledge of all possible features that a lexical item might have. With these features in mind, the algorithm is allowed to run, and the same structure should be built for any two lexical items that have the same feature sets. There is no allowance for the fact that lexical items that have the same features might nevertheless have different strengths for the different features. The experiments in this chapter show clearly that even though two lexical items allow the same sorts of complements, they will not necessarily behave the same. How might these results be accounted for within the confines of the SPARSE parsing system?

As alluded to at the end of Chapter 2, the most straightforward way to account for lexical biases is to assume that the lexical biases in some way restrict the features visible to the parser. For example, if a verb is equally biased between NP complements and sentential complements, both possibilities might always be available to the parser. However, for verbs that occur with sentential complements 75% of the time and with NP complements 25% of the time, the NP-complement features might not be available to the parser some portion of the time (perhaps 25% of the time). Under this account, an NP directly following the verb should always be analyzed as a direct object, by virtue of the fact that the direct object features are available. In contrast, NPs following S-biased verbs should be analyzed as direct objects only 75% of the time. The other 25% of the time the direct object features will not be available, and the NP will be analyzed as subject of a sentential complement. Thus, in conditions such as those in the experiments discussed above, the equi-biased verb should behave just like the NP-biased verbs, always showing ambiguity effects (which is taken to indicate misanalysis) in the low attachment conditions, and never showing evidence of misanalysis in high attachment conditions. 75% S-biased verbs should show ambiguity effects in both conditions. If it were possible to determine which analysis was made in each trial, roughly 75% of the trials should show a high attachment, and 25% should show a low attachment. This would predict that, all else being equal, the ambiguity effect in the low attachments (averaged over trials) should be much stronger than the ambiguity effect in the high attachments, because the high attachment cases would only require reanalysis in the 25%

of trials when the low attachment is made, while the low attachment disambiguations would require reanalysis in 75% of trials. However, in order to make this prediction, it is necessary to assume that the cost to reanalyze from high to low attachment is the same as the cost to reanalyze from low to high. Because the reanalyses themselves are different, I believe that it is most reasonable to assume that the cost of repair also differs between the two reanalyses. Because the difference in reanalysis cost is not known, it is difficult to make predictions about how the average cost of reanalysis should be related to verb biases.

A different formulation of what is basically the same account as in the previous paragraph is to assume that the features of a head vary in direct proportion to the frequency with which the feature occurs. In other words, if a verb occurs with an NP complement 50% of the time and with an S complement 50% of the time, the head will be seen by the parser as unambiguously taking an NP complement 50% of the time and will be seen as unambiguously taking an S complement 50% of the time. This formulation of the account makes different predictions than the account presented in the preceding paragraph. In particular, it predicts that equi-biased verbs should show ambiguity effects in both conditions, because the verb will be seen as rigidly taking one complement or the other at all times. The parser will pursue the high analysis of the verb 50% of the time (when the verb takes an NP complement) and will pursue the low analysis of the verb the other 50% of the time. Thus, ambiguity effects should be seen in both the high and the low conditions, because both options will be pursued.

Another question raised by these experiments is how to account for differences in resources available to subjects. The experiments showed that reading span scores are related to differences in parsing behavior. The parsing algorithm in Chapter 2 and Chapter 3 makes no reference to how these sorts of differences might be accounted for. However, as discussed above in ⁄ 5.4, there are a number of ways in which these differences might be related to the parsing algorithm. One option is to assume that high attachments are preferred when the parser is near the limits of its capacity. High attachments allow the parser to close off more structure than low attachments, and could therefore free up resources. This means that it should be easier to force low-span readers to make high attachments, because it will be easier to push them close to their limits. While this strategy accounts for the results from the experiments, it does not appear to be easily compatible with the parsing algorithm proposed here. In particular, this solution requires that the parser be able to fundamentally change its search process, from one that searches first for local attachments to one that first searches for non-local attachments. While this proposal might turn out to be correct, it seems more likely to me that the changes in parsing brought about by resource limitations are the result of something that is more directly related to resource limitations.

A proposal in which resource limitations are more straightforwardly related to changes in parsing behavior is that the differences in reading span are related to the ability of the parser to build and maintain non-overt heads (either phonetically null heads or predicted heads). If building new (null) nodes is associated with increased cost, attachments that involve fewer additional (null) nodes might be considered earlier in the search than would otherwise be the case. For example, because sentential complement analyses involve positing more nodes than direct object analyses require, a lexically disfavored direct object analysis might be preferred to the sentential complement analysis. An implementation of this idea might involve an early termination of an attachment search if the number of (null) nodes surpasses a certain limit. Thus, if a sententential complement analysis fails because there are too many null nodes, a simple direct object attachment might be discovered, even though it is lexically dispreferred. The details of how costs on the process of positing and maintaining new nodes interacts with lexical biases will require more research.

Chapter 6

CONCLUSIONS AND FUTURE WORK

6.1 Summary

The main goal of this work has been to provide a detailed account of how parsing human parsing can be modeled. I take word-by-word incrementality to be a hallmark of human parsing, and therefore a central criterion for determining whether or not a parsing model correctly accounts for human parsing behavior. While there is abundant evidence that English is parsed incrementally, there is relatively little experimental evidence from head-final languages demonstrating structure-dependent interpretations of complements in advance of the heads that take them as complements. This state of affairs is quite common in head-final languages, and I have attempted to show that there is good reason to believe that parsing in head-final languages proceeds just as incrementally as it does in head-initial languages. This evidence includes the fact that speakers seem to have available to them (before the structure-defining heads have been seen) the possible interpretations of a reflexive, which are crucially structure-dependent.

Accordingly, the discussion in the preceding chapters has concentrated on providing an explicit account of exactly how syntactic structures can be built incrementally in both head-initial and head-final languages. Crucial to the incremental processing of both types of languages is the use of predicted heads heads which will definitely occur in the sentence at some point, but which must be posited in advance of their occurrence in the sentence. In order to process sentences incrementally, some structure must be predicted. I have attempted to reduce the number of situations in which the parser might make overly strong predictions about upcoming material. When overly strong predictions are made in the course of a parse, they may later need to be retracted, and I assume that the fact that sentence processing generally proceeds very smoothly is evidence that the human parser is not engaged in extensive retraction of overly strong predictions. While I have discussed the use of predicted heads in the context of a serial parser, they are also potentially very useful within the limited parallel approach to parsing. In particular, it is clear that the parser can only carry a certain number of parses forward (though how many and which parses can be carried forward is very much in dispute). Therefore, a method for reducing the number of potential parses should be welcome. The use of predicted heads allows for a very significant reduction in the number of potential parses, because it allows a number of different parses to be collapsed into one.

The approach to predicted heads used in this thesis allows for significant underspecification of the features of the predicted heads. Because the predicted heads are only specified for the features that they will definitely contain, there is no need to decide in advance which other features the head will bear. For example, if a dative noun phrase is encountered in a head-final construction, there is no need to decide immediately whether the NP is a complement of a postposition, a dative verb, or a ditransitive verb. In a parallel parser, this has the advantage of reducing the number of potential parses from three to one, and in a serial parser it has the advantage that all three types of licensing heads can be accommodated without the need to retract any predictions.

As previously stated, the primary objective of this work has been to specify how structures can be found and built incrementally. However, I have also attempted to show how a number of parsing preferences can also be integrated into the parsing algorithm. Minimal Attachment effects, whereby attachments that require fewer additional nodes are preferred over attachments that require more nodes, are expressed in the algorithm by the way the search for an attachment is conducted, although the effect of Minimal Attachment in this model is different than in most other models because of the interaction with lexical preferences. Late Closure/Right Attachment/Recency effects are also tightly integrated into the algorithm. The main experimental finding is that attachments are preferred in positions that are structurally close to the most recently processed word.

This is expressed in the algorithm in the search process—the search for an attachment site proceeds along the right edge of the tree from the most recently processed word, in a manner that checks the nodes in order of their distance from the most recently processed word. Another finding implemented in the algorithm is the preference for argument attachments over adjunct attachments. This is straightforwardly built into the model by having the algorithm search for argument attachments before adjunct attachments are searched for. Specific interactions between these preferences are also predicted. For example, non-local argument attachments are predicted to be preferred over more local adjunct attachments, and recent attachments that require fewer additional nodes are predicted to be preferred over more recent attachments that require more additional nodes.[42]

This model also predicts that certain situations will be beyond the capabilities of the parser, resulting in parsing breakdown. Among the situations when this is predicted are situations in which a piece of structure is unavailable that is required to allow a grammatical attachment. Also predicted to cause breakdown are situations in which no single maximal projection can provide all and only the structure necessary to allow an attachment (i.e. if the required pieces of structure are scattered in several maximal projections or the maximal projection that contains the required structure also contains other structure).

An additional question about the point in the search process when reanalysis is considered was addressed experimentally. Using the word-by-word reading paradigm, experiments were conducted to determine whether local, easy reanalyses are preferred over less local analyses that do not involve reanalysis. The result is that the less local analysis is preferred over the local reanalysis. This result is somewhat surprising in light of the strong locality preferences that have been shown elsewhere in the literature. These results showed the answer for a fairly fundamental question about reanalysis when it is considered in the parse process. The clear answer is that reanalysis is only attempted as a

[42] Phillips and Gibson (1997a,b) provide results that appear to be at odds with this prediction.

last resort, after all other options for attachment have been exhausted. In accordance with this finding, reanalysis is a last resort option in the parsing algorithm presented here. The fact that locality is generally an important factor in parsing, but seems to be overridden in this particular instance was shown to be problematic for theories in which syntactic structure is built monotonically. These experiments also showed another apparently novel result that reflexives can be used as effective disambiguators is on-line experimentation.

In contrast to incremental parsing theories that make critical use of top-down structure-building (e.g. Stabler 1994, Gorrell 1995), structure-building in this theory is strictly bottom-up. There are two main motivations for the strictly bottom-up nature of the system.[43] First, I assume that the driving force behind parsing is the desire to interpret language. In order to interpret language, structures must be built that incorporate all of the input that has been received. As soon as the input has been received, interpretation may proceed. However, interpretation of the input does not require that any extra nodes be posited top-down. Any syntactic nodes that might be predicted top-down would contribute nothing to the interpretation of the input. The important goal for the parsing system is to structure all the words that have been received in the input. As long as those words have been structured, they can be interpreted and there is no need for any extra nodes. In this sense, I assume that the parser is basically lazy it does what is required for incremental structure-building and no more.

The second motivation for a purely bottom up parser is that it will be less prone to making mistakes. Given the great amount of ambiguity present in language, many nodes that might be predicted top-down would need to be retracted. As noted earlier, one of the goals of this work is to show a way to reduce the number of required reanalyses of

---

[43] Note that information about verb-biases does not trigger any top-down structure-building in this system. Instead, it serves only to further specify the combinatory possibilities of the verb. The fact that the verb bias indicates that an ambiguous verb can only use one of it combinatory possibilities makes that verb similar to a verb that only has one combinatory possibility to start with (i.e. an unambiguous verb).

syntactic structure. In light of this goal, the elimination of top-down predictions that frequently require retraction is a step in the right direction.

I have also shown how movement, a fairly fundamental part of language, can be accommodated within the incremental approach to parsing that is detailed here. The movement theory presented in Chapter 4, which is basically a translation of a traditional movement into a left-to-right incremental theory, accounts for a wide range of psycholinguistic and syntactic facts of wh-movement in a number of languages. The theory is in accord with experiments that show that there is an active effort to posit traces of movement in the positions where they are syntactically possible. The theory also accounts for the syntactic data showing that certain types of constituents are islands out of which constituents generally may not move. Movement out of islands in parasitic gap constructions is also shown to be compatible with the theory, by virtue of a mechanism very similar to the Principle of Minimal Compliance proposed in Richards (1997). Within the parsing theory presented here, this mechanism can also account for the fact that wh-islands are not respected in multiple wh-fronting languages like Bulgarian, and it is also compatible with parasitic movement out of other islands in Bulgarian. The account of wh-movement also shows that it is possible to build a representation for ungrammatical constructions, while still retaining knowledge of the ungrammatical status of a sentence.

6.2 Directions for Future Research

This work has focussed on the issue of incrementality, and has shown that it is possible to build a theory that accounts for the incremental nature of sentence processing. However, others have suggested that some structures are not processed incrementally. In future work I would like to experimentally investigate whether or not these structures are actually processed incrementally, as I would like to claim, or whether there are limits to the amount of incrementality that syntactic processing is subject to. For example, in the discussion of Categorial Grammar in ⁄1.2.2.4 , it was noted that the theory appears to predict that coordinated structures cannot be parsed incrementally. In particular, the second conjunct of a coordinate structure cannot be integrated into the structure with the first conjunct until the entire second conjunct is completed. Thus, CCG predicts that the reflexive *herself* in (240), repeated from (30), cannot be interpreted until the end of the sentence is reached.

(240) Dorothy killed the witch and deemed herself defender of the little people.

It has also been asserted by others that head-final structures are not parsed incrementally. However, there is very little experimental evidence showing that is true. Thus, this area would seem to be ripe for experimentation. One way that I would like to investigate this matter is through the use of reflexives (and possibly also pronouns). Under most syntactic theories, the interpretation of a reflexive is crucially dependent on the presence of syntactic structure. Thus, if reflexives are interpreted immediately, it would be clear evidence that syntactic structure is built incrementally. In ⁄3.1 , I introduced the sentences in (241) and (242) as the sort of sentences that might be used to show that structure is built incrementally.

(241) Die Frau glaubt, dass der Junge wegen seiner schlechten Noten sich selbst
the woman thinks that the boy because-of his bad grades him/her self
erschossen hat.
shot has
The woman thinks that the boy shot himself because of his bad grades.

(242) Der Mann glaubt, dass der Kollege wegen Geisteskrankheit ihn
The man thinks that the colleague$_{MASC}$ because-of mental-illness him
erschossen hat.
shot has
The man$_i$ thinks that the colleague$_j$ shot him$_i$ because he was mentally ill.

While the expected results from these sentences (that the reflexive can be interpreted immediately) would point in the direction of incremental parsing of these head-final constructions, the results would still be open to criticism, including the possibility that the complementizer might provide enough information to determine that the reflexive must be bound by an NP above the complementizer. Thus, it would be worthwhile investigating what sort of structures could overcome this objection.

Experimental work is also indicated for the various structures that can complete a temporarily ambiguous set of arguments in head-final languages, of the sort found in (96)-(98) and (101)-(108) in ⁄ 3.2. It is fairly clear that none of these structures causes parsing breakdown, but it is not clear that they are all equally easy to process. On-line experiments with these sorts of structures would provide much more fine-grained data to tune the theory of how predicted heads are combined with their overt counterparts.

Not addressed in the model so far is the question of how movement other than wh-movement can be dealt with. Wh-movement is in some respects easier than other types of movement because there are clear markers for when it is taking place. Other types of movement, such as scrambling and topicalization, are not as clearly marked and may therefore be somewhat more difficult to process. Because licensing heads are built on the basis of the features that are evident on heads, movement of heads without overt morphological marking may cause licensing heads to be predicted in the wrong locations. These licensing heads will then presumably need to be retracted or moved when it becomes clear that scrambling or topicalization has taken place. This topic therefore clearly merits further attention.

In a completely different direction, the fact that reading span had an effect on the experiments discussed in Chapter 5 raises the question of how the parsing system interacts with other cognitive processes and abilities. Other psycholinguistic research (e.g. King and Just 1991) has shown that there is an interaction between certain types of cognitive abilities and performance in sentence processing. However, these experiments are typically done with experimental materials that are at the very edge of understandability. The thinking seems to be that most language processing falls within the range of most normal speakers, and that language processing is more or less equally efficient for all people, though the limits on their capabilities may differ. No interaction with other cognitive processes/systems was expected in these experiments, because the stimuli appear to be well within the grasp of normal subjects. The fact that an interaction was seen indicates that this model may not be correct. Instead, it may be the case that some people are simply better (e.g. faster) at processing language than others. The differences in the limits of speakers capacities may be caused by the difference in the basic abilities and speeds of the speakers. Whatever the reason for the difference in performance between the two groups studied in the experiments, a clear explanation of how the difference can be integrated into the model proposed here would be very useful.

Related to inter-speaker differences in cognitive and language processing abilities are the studies of people with language deficits. Models such as those presented by Cornell (1995), Haarmann and Kolk (1991), and Haarmann, Just, and Carpenter (1997) show that language deficits can be modeled by imposing constraints on standard models of language processing. The model of parsing that Haarmann and Kolk used was not an incremental model, and further research is needed to determine if the results obtained in that work can be accommodated within a more incremental parser, such as the one proposed here.

As far as computational matters are concerned, one outstanding question has to do with the theoretical efficiency of the model proposed here. It is not clear to me how one could even determine the theoretical complexity of the model proposed here, because of the way structures can be changed. More investigation into the computational properties of the structure-changing operations proposed here would certainly be warranted.

Finally, in order to more fully test the parsing theory presented here, the implementation of the theory (described in Appendix A) needs to be extended. At the present time, only part of the theory is implemented, making it difficult to determine whether more clarification is needed in the model. Likewise, it cannot be easily determined if there are structures that this theory is simply not capable of accounting for. One very useful improvement to the implementation would be a processing metric, which would allow the amount of processing required by the model to be compared with actual experimental results. Such a metric might, for example, count the number of times each step in the attachment algorithm is executed, with possibly differing weights for the different operations.

Appendix A

IMPLEMENTATION

A portion of the parser developed in the work above was implemented as a computer program. The parser was implemented in Java as a fully object-oriented application. There were two main rationales for implementing the parser: to ensure that the theory was implementable and to ensure that the theory was very explicit. As noted in Chapter 1, psycholinguistic parsing theories often cannot be fully evaluated because they are too vague. By implementing the theory discussed here, I have been forced to make the theory specific enough that it can be fully evaluated.

Java was chosen as the programming language for a number of reasons. One reason is that it is a fully object-oriented language. Linguistics is an area that is particularly well-suited to object orientation, since most of the symbols in linguistics make for natural object types. For parsing purposes, features, heads, and constituents are natural objects, and the entire program is built around these objects. Java, including its graphical user interface (GUI) functions, is also designed to be platform-independent, a desirable attribute for a program that is designed to be usable by a variety of researchers. Finally, the use of Java was also meant to be a good learning experience for the author.

Below, I describe the data structures used in the implementation, the parsing algorithm, the lexicon, and the user interface.

A.1 Data Structures

A.1.1 Features

The feature object serves as the basic building block for the entire parser. In the implementation, a feature consists of a feature name (e.g. Case), a set of values, a flag for storing whether or not a feature has been checked, and in the case of licensing features, a direction of assignment. In the present implementation, there is no facility for indicating whether licensing features are specifier, complement, or adjunct licensing features.

In addition to containing slots for data elements, feature objects also include a number of functions (called methods in Java) that operate on features. The most important method for features is the feature that determines whether or not two features can enter into a checking relation. This method first determines whether or not two features have the same feature name and whether either of the features has already been checked off. If they have the same name and have not been checked, an intersection is performed on the feature values. If the intersection shows that the two features have at least one value in common, the method returns true, indicating that the two features may enter into a checking relation. If either of the features has the value *variable* (indicating that the feature can take any value), the checking method always returns true. An additional checking method is available to take actually establish a checking relation between two features. This method changes the value of the features to the intersection of the values of the two features. If one of the features has the value *variable*, the value of the other feature is copied in place of *variable*. In addition to setting the values of the features, this method also sets the checked flag on both features to indicate that they have both entered into a checking relation and are therefore ineligible for further use in the parse.[44]

---

[44] Checked features can still be used in the parse, but only when the features are involved in reanalysis.

Features also contain methods used for lexicon construction. These methods allow things like adding feature values, setting the name of a feature, and similar utility functions.

A.1.2 Nodes

There are two main types of syntactic nodes in the implementation: heads and binary-branching nodes called multiconstits . In order to ensure that both type of nodes can be used interchangeably in syntactic operations, a node is defined. The node is a Java construct (technically speaking, an interface) that ensures that both heads and multiconstits contain the methods specified in the interface definition. Methods required by the node interface include methods that return a pointer to the parent of the node, methods to determine whether the node can license the attachment of another node (detailed in ⁄ A.2), and other similar methods.

A.1.2.1 Heads

Heads are the basic object manipulated by the parsing algorithm. Heads include the following data structures: a set of inherent features, a set of licensing features, the minimal head required to license the head, a specification of the word the node instantiates, and a flag indicating whether the head must be found in the input before it can be used by the parser, and a pointer to the head s parent node. The two sets of features correspond to the sets of features discussed above in Chapter 2. In the implementation, there is no notion of the distinguished feature that is used in the rest of this dissertation. Recall that the distinguished feature is used as the basis for construction of predicted heads in head-final languages. The work done by the distinguished feature is performed in the implementation by a specification of the minimal licensing head. The minimal licensing head is a specification of what features must be present on a predicted head in order to license the associated head in a syntactic structure. The fact that this set of features is specified directly on all heads requires that the information be compiled out ahead of time from all possible licensers of a word. In future work on the implementation, I plan to replace the minimal licensing head specification with a distinguished feature and the methods necessary to search the lexicon in the ways described above in Chapters 2 and 3.

In order to ease understanding of the trees built in the course of parsing, each head is specified for a word that it represents, even if the head is phonologically null (e.g. the null complementizer in English, which is called nullC). Because some heads are specified as representing a word even when that word is phonologically null (e.g. nullC), each head also has a flag indicating whether or not the head must be encountered in the input before it can be used. The details of how this flag is used can be found in / A.2 and /A.1.3.

Head objects contain all of the methods specified in the node interface, as well as a number of specialized methods that are not required for multiconstits. Among these are methods that are used to build the lexicon, such as methods to add features to heads, specify the word associated with the head, and make a copy of the head.

A.1.2.2 Multiconstits

The other type of syntactic constituent used in the implementation is the multiconstit. A multiconstit is a binary-branching node, the children of which can be either heads or other multiconstits. Because of the nature of bare phrase structure, multiconstits themselves bear no syntactic features. Multiconstits contain nothing more than pointers to the two children and the parent, and an indication of whether the head of the multiconstit is in the left or right child.

Multiconstits contain all of the methods required in the node interface, as well as methods to build multiconstits from two syntactic nodes. During parsing, requests are frequently made for the features of a given multiconstit. Because multiconstits contain no syntactic features, the values of the head of the constituent are returned instead. For this reason, multiconstits also contain methods to search down the tree for the head of a constituent. These methods follow the head-direction features on the multiconstits down the tree until they arrive at the head of the constituent, at which point the relevant information from the head of the constituent is returned.

A.1.3 Lexicon

The lexicon object stores heads and multiconstits (e.g. tensed verbs that are stored as a constituent containing both a verb and a Tense head). When a lexical item is encountered in the input, a search of lexicon is conducted and a copy of the head/multiconstit corresponding to the input word is returned. It is also possible to search the lexicon for items that contain specific features or are able to license a particular syntactic node. One search method searches only for heads that are not present in the input. This is useful for determining whether or not some null element (e.g. a null complementizer or determiner) to the left of the incoming word might be able to license it. As discussed above in ⁄ A.1.2.1, this implementation does not search the lexicon to determine what type of predicted head should be built. However, most of the methods necessary to do this are in place. Methods to perform a search of the lexicon for heads containing a particular feature and to perform feature intersection are already built. In order to fully implement lexically-based predictions, a method to intersect heads, as well as methods to trigger the search would be necessary.

At the present stage, lexical items must be built into the code. A lexical entry is built by building each of its features, adding their feature values one at a time, at which time each of the features can be added to the head. Once a full head is built, it can be added to the lexicon. Two lexicons are presently included: German and English. Each of them has approximately 20 lexical items, which is enough to allow a number of interesting sentence types to be investigated.

A.2 Parsing algorithm

The implemented version of the parsing algorithm is capable making simple attachments into the existing tree, and also contains facilities for building extra structure when a direct attachment into the existing tree is not possible. The search for a direct attachment starts at the most recent word in the input. The search involves checking for a licensing relation between the incoming word and the present node in the tree (which starts out as the most recently processed word). A licensing relation either from the new word to the existing node or from the existing node to the new word is enough to allow attachment and trigger processing of the next word in the input. If no licensing relation is found, the search continues up the tree until an attachment site is found, a predicted head is reached, or the root of the tree is encountered. If a licensing relation is found, the new word is made a sister of the existing node, and the parent node is specified for the appropriate headedness. Because the mechanism for reanalysis has not yet been implemented, the node in the existing tree is always the head of the new constituent (i.e. direct attachment always involves either a complement relation or an adjunct attaching to a head on its left).

If the root of the tree is reached, a predicted head is built to license the incoming word. The entire attachment process is then attempted for the new constituent (the incoming word and its predicted licenser). If a predicted head is reached in the search process, one of two things can happen. If the features of the predicted head can be subsumed by the incoming word (e.g. a predicted accusative assigner can be subsumed by an accusative verb), the predicted head and the instantiated head will be merged into one this amounts to instantiating the predicted head with the word from the input. If this happens, the processing of the word is complete and the processing of the next word is triggered. If the predicted head cannot be subsumed by the incoming word, the search for an attachment site for the new word is terminated, a predicted head is built to license the incoming word, and the entire attachment process is restarted for the new constituent. Subsumption of multiple heads by one word (e.g. a ditransitive verb) is not implemented.

As noted above, when no direct attachment/subsumption into the existing tree is possible, a new licensing head is attached to the incoming word. The lexicon of phonetically null elements is first searched, to see if any of the null elements is capable of licensing the existing word. If no null element is available, a predicted licensing head is built. The licensing head is built on the basis of the minimal licensing head that is stored in the lexical representation of each head. The licensing head is then attached to the new word. Once the extra head (either a null element or a predicted head) has been attached to the new word, the entire parsing process is run again with the new constituent functioning as the new word in the input. This new search for an attachment starts at the most recently processed word and proceeds up the tree in the normal fashion.

A.3 User Interface

The user interface consists of a graphical user interface in which the user can load the lexicon, input the sentence to parse, see the tree that is produced, and view the contents of the heads used in the parse. In Figure 28 below, the interface is shown.

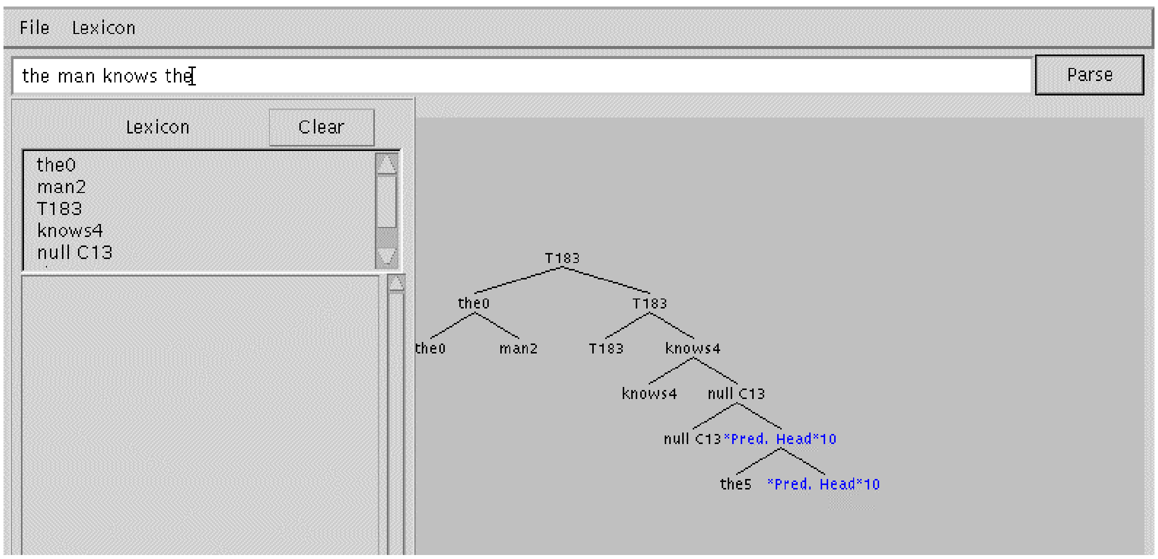


Figure 28: Interface state for *the man knows that* (note: *knows* is specified to only take a sentential complement in this example.)

As can be seen, the user is able to enter the sentence into a textbox, and can then proceed to parsing. It is also possible to put the parser in an incremental mode, in which one word of the input is processed each time the Parse button is pressed. The upper left window shows the heads that are used in the parse. Clicking on any one of them causes the entire contents of the head (i.e. the features and their values) to be output in the window below. Each word used in the parse has a unique serial number to identify the head to the user. This is done to differentiate the different instances of the same word. For example, in Figure 28, the word *the* is used twice. The serial numbers provide a way for the user to see the features for a particular instance of *the*, either *the0* or *the5*. Predicted heads are indicated with by the string *Pred. Head* and are additionally highlighted by being shown in blue on the screen.

Items in the lexicon menu allow the user to load the German and English lexicons, and they also allow the user to view the lexicon items directly in the head-viewing window.

Appendix B

EXPERIMENTAL MATERIALS

B.1 Experiment 1 Stimuli

The sentences used in Experiment 1 followed the schema outlined in (246)-(249) below. The underlined materials indicates the differences between the conditions. In the two ambiguous conditions, the stimuli are exactly the same up to the reflexive. The high unambiguous conditions were disambiguated by the use of an accusative pronoun (*her* in (247)) instead of a full, definite NP after the embedding verb. The low unambiguous conditions included the complementizer *that* immediately after the embedding verb to force the embedded subject reading of the following NP. Following the reflexive in each condition was an identical four-word PP. After the PP, the low conditions included another verb to make the stimuli grammatical, while in the high conditions, other extra material was included to keep any end-of-sentence effects away from the critical regions.

(246) *High Ambiguous*
The smart fellow who mentioned the senator s wife got a job for himself with the powerful lobbyist soon thereafter.

(247) *High Unambiguous*
The smart fellow who mentioned her got a job for himself with the powerful lobbyist soon thereafter.

(248) *Low Ambiguous*
The smart fellow who mentioned the senator s wife got a job for herself with the powerful lobbyist got some hush money.

(249) *Low Unambiguous*
The smart fellow who mentioned that the senator s wife got a job for herself with the powerful lobbyist got some hush money.

This set of stimuli corresponds to sentence 1 below.

MENTION

1. The smart fellow who mentioned (that) {the senator s wife/her} got a job for {himself/herself} with the powerful lobbyist {soon thereafter/got some hush money}.

2. The quiet woman who mentioned (that) {the violent man/him} caused big trouble for {herself/himself} with the local police {by discussing the incident at all/was thought to be an informant}.

3. The busy woman who mentioned (that) {her sick son/him} picked up the prescription {herself/himself} from the neighborhood pharmacy {after work/was accused of neglect}.

4. The nervous man who mentioned (that) {the accused woman/her} cursed very loudly at {himself/herself} before the important hearing {about the robbery/was asked about the robbery}.

5. The attentive waitress who mentioned (that) {the sick man/him} opened the bathroom door {herself/himself} in a big hurry {so he could get in/sent someone to see if he needed help}.

DOUBT

6. The experienced monk who doubts (that) {the new nun/her} wrote in the ledger {himself/herself} about the membership increase {of the last five years/told the bishop about the potential problem}.

7. The angry woman who doubts (that) {the lazy man/him} called the state police {herself/himself} at about three o clock {to report her suspicions/thinks his wife made the call}.

8. The pig-headed girl who doubts (that) {the stupid boy/him} reported the transgressions {herself/himself} to the principal {after school/thinks the class snitch was the culprit}.

9. The distrustful woman who doubts (that) {the strange man/him} locked the front door {herself/himself} after the big party {for the neighbors/asked who really locked the door}.

10. The famous knight who doubts (that) {the lazy girl/her} fed the hungry horses {himself/herself} before the evening meal {of potatoes and roast beef/looked for footprints in the dirt}.

NOTICE

11. The handsome man who noticed (that) {the famous actress/her} called the fashion magazine {himself/herself} after the gala fundraiser {for the charity/was surprised the woman s assistant didn t call}.

12. The nice man who noticed (that) {the sick woman/him} opened the bathroom door {himself/herself} despite the strong protestations {of the staff/ of the staff tried to reassure her}.

13. The observant woman who noticed (that) {the mistreated man/him} informed the personnel department {herself/himself} about the frightening incident {a few days after it happened/was pleased with the response}.

14. The young mother who noticed (that) {the disabled man/him} warned the hospital staff {herself/himself} about the broken wheelchair {as soon as she could/felt sorry for him}.

KNOW

15. The middle-aged man who knows (that) {the famous woman/her} invited the political press {himself/herself} to the news conference {about the big scandal/was not pleased by her action}.

16. The smart girl who knows (that) {the foolish boy/him} told the long story {herself/himself} during the story hour {at the library/gained a new respect for him}.

17. The creative woman who knows (that) {the funny man/him} wrote some comedy sketches {herself/himself} about the amusing escapades {she had seen/thinks he should publish them}.

18. The bilingual man who knows (that) {the well-traveled woman/her} translated the travel books {himself/herself} without any extra help {from a dictionary/was quite impressed with the result}.

19. The nosy guy who knows (that) {the depressed woman/her} told some hilarious jokes {himself/herself} at the doctor s office {in order to cheer her up/thinks she might be faking her depression}.

FEAR

20. The paranoid man who fears (that) {the deceptive woman/her} locks the barroom doors {himself/herself} after the last call {every night/thinks he will be locked in}.

21. The grumpy woman who fears (that) {the scruffy man/him} called the building management {herself/himself} after the late-night disturbance {in the stairwell/tried to appease him}.

22. The anxious man who fears (that) {the deranged woman/her} wrote the accusatory statement {himself/herself} at the police station {to get her arrested/expects the police to visit}.

23. The sleepless woman who fears (that) {the night watchman/him} checks the window blinds {herself/himself} during the late movie {every night/is planning to buy a dog}.

24. The inexperienced nanny who fears (that) {the cranky handyman/him} fixed the squeaky door {herself/himself} with some cooking oil {so the man wouldn t have to come in the house/knows the door will need to be fixed again soon}.

HEAR

25. The quiet man who heard (that) {the terrified woman/her} called the county police {himself/herself} at around four o clock {in the morning/in the morning was shocked at the lack of any response}.

26. The quiet girl who heard (that) {the mischievous boy/him} opened the front door {herself/himself} despite many strong warnings {that the boy shouldn t be allowed in/hoped he wouldn t get in trouble}.

27. The intelligent man who heard (that) {the clumsy woman/her} turned on the light {himself/herself} on the front porch {so she wouldn t fall/decided to install an automatic light}.

28. The caring woman who heard (that) {the depressed man/her} called the mental hospital {herself/himself} about the serious side-effects {he was experiencing/was relieved at the news}.

DISCOVER

29. The kind man who discovered (that) {the injured woman/her} called the emergency room {himself/herself} from a nearby payphone {after the late-night accident/offered to drive her to the hospital}.

30. The surprised woman who discovered (that) {the drunk man/him} locked the front door {herself/himself} with the spare key {to keep him away/was amazed he could even walk}.

31. The hungry man who discovered (that) {the starving woman/her} stole the big roast {himself/herself} from the deserted kitchen {of the restaurant/asked her to share some with him}.

32. The angry woman who discovered (that) {the badly-beaten man/him} protested to the authorities {herself/himself} about the rampant violence {for over an hour/ was proud of his courage}.

UNDERSTAND

33. The fickle actress who understood (that) {the strange man/him} wrote the bizarre stories {herself/himself} in just three days {despite being sick/wondered at the man s brilliance}.

34. The generous nobleman who understood (that) {the persuasive woman/her} bought the powerful telescope {himself/herself} for the small college {in order to promote science/promised to fund a new observatory}.

35. The pious man who understood (that) {the religious woman/her} financed the church restoration {himself/herself} with the large inheritance {from a distant cousin/was impressed with her devotion}.

36. The unhappy man who understood (that) {the shy woman/her} reported the continuing problems {himself/herself} in an angry phone message {to the apartment manager/visited the woman to settled the dispute}.

ACKNOWLEDGE

37. The kind man who acknowledged (that) {the considerate woman/her} typed the generous contribution {himself/herself} into the charity database {after receiving the check/showed appreciation for the help}.

38. The magnanimous prince who acknowledged (that) {the quiet woman/she} accepted all the blame {himself/herself} for the many errors {of his subjects/pardoned her after the investigation}.

39. The grateful woman who acknowledged (that) {the maintenance man/him} answered all the questions {herself/himself} about the serious incident {at the refinery/thanked him profusely}.

40. The beloved duchess who acknowledged (that) {the rich man/him} saved the well-known charity {herself/himself} despite the financial problems {following the scandal/praised him at the banquet}.

WARN

41. The big policeman who warned (that) {the angry saleswoman/her} ran down the street {himself/herself} after the getaway car {following the robbery/hoped the other cops wouldn t shoot her}.

42. The neighbor woman who warned (that) {the macho boy/him} shot the rabid dog {herself/himself} with a large shotgun {before going to work/thought she should call his parents}.

43. The concerned priest who warned (that) {the meddling woman/her} opened all the mail {himself/herself} in the church office {so she wouldn t read the letters/told the worried parishioner to contact him by phone}.

44. The young widow who warned (that) {the little boy/him} opened the locked cabinet {herself/himself} with a small screwdriver {while he was outside/told the babysitter to watch him closely}.

APPRECIATE

45. The health-conscious man who appreciates (that) {the beautiful woman/her} takes good care of {himself/herself} throughout the entire year {by eating regularly and walking everywhere/wants to ask her on a date}.

46. The ambitious boy who appreciates (that) {his smart sister/her} does lots of homework {himself/herself} after swim team practice {because he wants to get into a good college/understands that she doesn t have time to help him}.

47. The good-natured landlady who appreciates (that) {the nice man/him} cleans the big house {herself/himself} before the day s end {so the house will be clean for him/is thinking about lowering his rent}.

48. The thoughtful woman who appreciates (that) {the young man/him} pays all the utilities {herself/himself} before the due date {to keep the man happy/wants to keep the man as a roommate}.

B.2 Experiment 2 Stimuli

The stimuli in Experiment 2 are shown below in a manner like that of Experiment 1. The full set of stimuli corresponding to stimulus 1 below can be seen in (250)-(253).

(250) *High Ambiguous*
The congressional staffers were not surprised that the smart fellow who mentioned the senator s wife leaked the important news himself to the powerful lobbyist soon thereafter.

(251) *High Unambiguous*
The congressional staffers were not surprised that the smart fellow who mentioned her leaked the important news himself to the powerful lobbyist soon thereafter.

(252) The congressional staffers were not surprised that the smart fellow who mentioned the senator s wife leaked the important news herself to the powerful lobbyist got some hush money.

(253) The congressional staffers were not surprised that the smart fellow who mentioned she leaked the important news herself to the powerful lobbyist got some hush money.

As noted in Chapter 5, the stimuli were split into two different blocks. Stimuli 31-40 in each block contain only the high conditions, because the embedding verbs in those stimuli do not allow sentential complements, but rather only allow direct object NP complements.

B.2.1 Block A

MENTION

1. The congressional staffers were not surprised that the smart fellow who mentioned {the senator s wife/her/she} leaked the important news {himself/herself} to the powerful lobbyist {soon thereafter/got some hush money}.

2. The neighborhood gossips reported that the busy woman who mentioned {the sick boy/him/he} picked up the prescription {herself/himself} from the local pharmacy {after work/was accused of neglect}.

3. The restaurant managers were glad that the attentive waitress who mentioned {the sick man/him/he} opened the bathroom door {herself/himself} in a big hurry {so he could get in/sent someone to see if he needed help}.

DOUBT

4. The family friends stated that the angry woman who doubts {the apathetic man/him/he} called the state police {herself/himself} at about three o clock {to report her suspicions/thinks the man s son made the call}.

5. The curious children observed that the anxious woman who doubted {the strange man/him/he} locked the front door {herself/himself} after the big party {for the neighbors/asked who really locked the door}.

NOTICE

6. The editors proved that the handsome man who noticed {the famous actress/her/she} called the fashion magazine {himself/herself} after the gala fundraiser {for the charity/tried to interrupt the call}.

7. The witnesses discovered that the conscientious woman who noticed {the mistreated man/him/he} informed the personnel department {herself/himself} about the frightening incident {a few days after it happened/had fired several employees}.

8. The rowdy boys remarked that the thoughtful man who noticed {the young girl/her/she} closed the fireplace doors {himself/herself} before the birthday party {at the mansion/congratulated her on her forethought}.

KNOW

9. The proud parents boasted that the smart girl who knows {the childish boy/him/he} told the long story {herself/himself} during the story hour {at the library/gained a new respect for him}.

10. The book publisher heard that the bilingual man who knows {the well-traveled woman/her/she} translated the travel books {himself/herself} without any extra help {from a dictionary/was quite impressed with the result}.

HEAR

11. The neighbors were surprised that the quiet man who heard {the terrified woman/her/she} called the county police {himself/herself} at around four o clock {in the morning/in the morning didn t say anything to anybody}.

12. The family members believe that the intelligent man who heard {the handicapped woman/her/she} turned on the light {himself/herself} on the front porch {so the woman wouldn t fall/decided to install an automatic light}.

DISCOVER

13. The hospital staff was surprised that the kind man who discovered {the injured woman/her/she} called the emergency room {himself/herself} from a nearby payphone {after the late-night accident/offered to drive her to the hospital}.

14. The vagrants stated that the hungry man who discovered {the starving woman/her/she} stole the big roast {himself/herself} from the deserted kitchen {of the restaurant/asked her to share some with him}.

UNDERSTAND

15. The movie studio was surprised that the fickle actress who understood {the strange man/him/he} wrote the bizarre stories {herself/himself} in just three days {despite being sick/wanted to put on the stories}.

16. The church leaders thought that the pious man who understood {the religious woman/her/she} financed the church restoration {himself/herself} with the large inheritance {from a distant cousin/was impressed with her devotion}.

ACKNOWLEDGE

17. The institute employees believe that the kind man who acknowledged {the considerate woman/her/she} typed the generous contribution {himself/herself} into the charity database {after receiving the check/mentioned her to the board members}.

18. The corporate office was pleased that the grateful woman who acknowledged {the maintenance man/him/he} answered all the questions {herself/himself} about the serious incident {at the refinery/thanked him profusely}.

WARN

19. The store managers thought that the big policeman who warned {the angry saleswoman/her/she} ran down the street {himself/herself} after the getaway car {immediately after the robbery/told the other cops not to shoot}.

20. The nuns thought that the concerned priest who warned {the meddling woman/her/she} opened all the mail {himself/herself} in the church office {so she wouldn t read the letters/told the worried parishioner to contact him by phone}.

SUSPECT

21. The truancy office knows that the concerned woman who suspected {the short boy/him/he} called the youth center {herself/himself} after the recent fight {to report what she knew/intends to talk to the boy s parents}.

22. The executive committee thinks that the perceptive woman who suspects {the friendly man/him/he} started the malicious rumor {herself/himself} during the civic meeting {because she wanted to keep the man from getting elected/didn t tell anyone of her suspicions}.

23. Some classmates reported that the scheming girl who suspects {the grumpy boy/him/he} ate the incriminating evidence {herself/himself} before the school meeting {so the boy couldn t get her in trouble/told the teacher about her suspicions}.

24. The loan company learned that the depressed man who suspects {the devious woman/her/she} sold the engagement ring {himself/herself} at the pawn shop {for five-hundred dollars/went to the pawn shop to look for the ring}.

BELIEVE

25. The coaches are amused that the young boy who believes {the neighbor girl/her/she} buys some peppermint gum {himself/herself} before every baseball game {because the gum is supposed to bring the team good luck/hopes she will give him some gum}.

26. The dealership knows that the dark-haired woman who believes {the knowledgeable salesman/him/he} wants the blue car {herself/himself} despite the peeling paint {on the car/the car is a good deal}.

27. The police stated that the charming woman who believes {the dishonest man/him/he} bought the valuable artifacts {herself/himself} from the disreputable store {at the man s insistence/thinks the man deserves to be cheated}.

CLAIM
28. The agency staff was surprised that the excited woman who claimed {the small boy/him/he} opened the front door {herself/himself} without any help {despite holding the boy with both hands/made such a big deal about it}.

29. The sales clerks couldn t believe that the kind woman who claimed {the errant boy/him/he} purchased the expensive clothing {herself/himself} with a credit card {so that the store would not prosecute him for theft/doesn t think the boy s father will believe her}.

30. The dispatch office stated that the dismayed mother who claimed {the runaway boy/him/he} called the state police {herself/himself} from a pay phone {because the phone at home was broken/thinks the police should have gone to the pay phone to pick the boy up}.

SUPPORT
31. The homeless people think that the brilliant woman who supported {the impoverished man/him} called the large charity herself after the fund drive to see if he had applied for aid.

FIRE
32. The office supervisors know that the nice man who fired {the incompetent woman/her} wrote the critical evaluation himself during the long flight back to the home office.

PROSECUTE
33. The smuggling ring fears that the stern woman who prosecuted {the violent man/him} called the immigration office herself after the deportation hearing to make sure that he couldn t get back in.

DEFY
34. The alert siblings remarked that the vengeful boy who defied {the young woman/her} opened the large package himself on the kitchen table.

RESCUE
35. The concerned relatives were glad that the valiant woman who rescued {the injured man/him} battled the car fire herself for over fifteen minutes until the firefighters arrived.

treat

36. The office staff said that the compassionate man who treated {the depressed woman/her} filed the insurance forms himself with the insurance company the woman had.

ADORE
37. The many bystanders thought that the concerned woman who adores {the teenage boy/him} called the rescue squad herself on a cell phone in her purse.

ANNOY
38. The delivery staff said that the well-known man who annoyed {the quick-tempered woman/her} sent the accusatory letter himself with a polite note taped to the top.

ABUSE
39. The police officers claimed that the cruel man who abused {the penniless woman/her} locked the thick door himself with the monstrous key so that she couldn t get out.

PITY

40. The children s teachers thought that the sweet girl who pitied {the poor boy/him} bought the new jacket herself as a birthday gift for the boy.

B.2.2 Block B

MENTION

1. The private detectives think that the quiet woman who mentioned {the violent man/him/he} caused big trouble for {herself/himself} with the local police {by discussing the incident at all/is a big gossip}.

2. The court officials were amazed that the nervous man who mentioned {the accused woman/her/she} cursed very loudly at {himself/herself} before the important hearing {about the robbery/was asked about the robbery}.

DOUBT

3. The concerned parishioners heard that the experienced monk who doubts {the new nun/her/she} wrote in the ledger {himself/herself} about the attendance increase {of the last five years/told the bishop about the potential problem}.

4. The nosy classmates assumed that the pig-headed girl who doubts {the stupid boy/him/he} reported the transgressions {herself/himself} to the principal {after school/thinks the class snitch was the culprit}.

5. The talkative servants suspected that the famous knight who doubts {the lazy girl/him/he} fed the hungry horses {himself/herself} before the evening meal {of potatoes and roast beef/looked for footprints in the dirt}.

NOTICE

6. The anxious family appreciated that the nice man who noticed {the sick woman/her/she} opened the bathroom door {himself/herself} despite the strong protestations {of the staff/of the staff tried to reassure her}.

7. The visitors said that the young mother who noticed {the disabled man/him/he} warned the hospital staff {herself/himself} about the broken wheelchair {as soon as she could/put it in the corner}.

KNOW

8. The persistent reporters remarked that the middle-aged man who knows {the famous woman/her/she} invited the political press {himself/herself} to the news conference {about the big scandal/was not pleased by her action}.

9. The talent agency thinks that the creative woman who knows {the funny man/him/he} wrote some comedy sketches {herself/himself} about the amusing escapades {she had seen/ wants to publish them}.

10. The office staff learned that the nosy guy who knows {the depressed woman/her/she} told some hilarious jokes {himself/herself} at the doctor s office {in order to cheer her up/thinks she might be hiding her depression}.

HEAR

11. The anxious friends were glad that the quiet girl who heard {the mischievous boy/him/he} opened the front door {herself/himself} despite many strong warnings {to leave the door locked/doesn t plan to tell anyone}.

12. The co-workers reported that the caring woman who heard {the depressed man/him/he} called the mental hospital {herself/himself} about the serious side-effects {he was experiencing/was relieved to hear the news}.

DISCOVER

13. The freezing friends were upset that the surprised woman who discovered {the drunk man/him/he} locked the front door {herself/himself} with the spare key {to keep him away/took the key from him}.

14. The concerned relatives heard that the angry woman who discovered {the badly-beaten man/him/he} protested to the authorities {herself/himself} about the rampant violence {for over an hour/was proud of the man s courage}.

UNDERSTAND

15. The faculty members were grateful that the generous benefactor who understood {the persuasive woman/her/she} bought the powerful telescope {himself/herself} for the small college {in order to promote science/promised to fund a new observatory}.

16. The irritated neighbors noted that the unhappy man who understood {the shy woman/her/she} reported the continuing problems {himself/herself} in an angry phone message {to the apartment manager/ visited the woman to settle the dispute}.

ACKNOWLEDGE

17. The anxious family was delighted that the smart prince who acknowledged {the quiet woman/her/she} accepted all the blame {himself/herself} for the judgement errors {of his subjects/pardoned her after the investigation}.

18. The board members were gratified that the beloved duchess who acknowledged {the rich man/him/he} saved the popular charity {herself/himself} despite the financial problems {following the scandal/praised him at the banquet}.

WARN

19. The teenage friends feared that the neighbor woman who warned {the aggressive boy/him/he} shot the wild dog {herself/himself} with the powerful rifle {before going to work/would call the boy s parents}.

20. The visiting children thought that the young widow who warned {the little boy/him/he} opened the locked cabinet {herself/himself} with a small screwdriver {while the boy was outside/told the babysitter to watch him closely}.

SUSPECT

21. The newspaper said that the scared man who suspects {the irate woman/she/her} notified the police department {himself/herself} after the recent burglary {at the neighborhood store where the woman works/hopes that the woman will be arrested}.

22. The supervisors realized that the astute man who suspected {the brilliant woman/she/her} made the amazing discovery {himself/herself} in the government lab {before the woman announced her results/thinks that she should have taken the credit}.

23. The police think that the ailing woman who suspects {the ill-tempered boy/him/he} took the cash box {herself/himself} from the unlocked cabinet {so that the boy couldn t steal it/told the department manager what she thought had happened}.

BELIEVE

24. The foundation employees heard that the destitute man who believes {the kind woman/her/she} visited the soup kitchen {himself/herself} after the severe storm {to get a good meal/asked about her health}.

25. The court jesters joked that the beautiful princess who believes {the dim-witted knight/him/he} contacts the war committee {herself/himself} before every committee meeting {to try to get the knight promoted/should worry more about someone else}.

26. The neighbors mentioned that the desperate man who believes {the cheerful woman/her/she} bought the new car {himself/herself} at the car dealership {that {the woman/she} recommended/ thinks she made a foolish decision}.

27. The office staff wasn t surprised that the insecure man who believes {the reassuring woman/her/she} opened the mysterious package {himself/herself} before the long flight {because she told him to look inside/called to see if she liked the gift}.

CLAIM

28. The security company thought that the concerned father who claimed {the screaming girl/her/she} broke the small window {himself/herself} with a small rock {so he could open the door to rescue {his daughter/her}/said so to stay out of trouble}.

29. The family said that the aging uncle who claimed {the teenage girl/her/she} purchased the new car {himself/herself} at the car dealership {so {the girl/she} could have his old car/denied that he had anything to do with the purchase}.

30. The security officers stated that the scheming man who claimed {the little girl/her/she} removed the name tag {himself/herself} with a quick jerk {so no one would know that {the girl/she} was not his/ tried to take {the girl/her} from the hospital}.

HUMILIATE

31. The office workers remarked that the arrogant man who humiliated {the sad woman/her} paid the large fee himself with a personal check to show how rich he was.

LOVE

32. The family members heard that the intelligent woman who loves {the good-looking man/him} called the fancy restaurant herself for dinner reservations that evening.

33. The school classmates guessed that the smiling father who loves {the precocious girl/her} bought the expensive doll himself as a birthday present for the girl.

ADMIRE

34. The hotel staff noticed that the thoughtful man who admires {the ambitious woman/her} opened the front door himself at the sea-side resort just in time for the woman to enter.

DISAPPOINT

35. The sympathetic lawyers were disappointed that the devious woman who deceived {the stupid man/him} claimed the large reward herself from the insurance company after the trial.

DESPISE

36. The staff supervisors remarked that the cruel man who despises {the stupid woman/her} presented the important report himself at the afternoon meeting so he could claim credit for it.

HARASS

37. The family members noticed that the grumpy woman who harassed {the lazy man/him} stained the new couch herself with a greasy hamburger just after the couch was delivered.

MEDICATE

38. The nursing staff said that the conscientious man who fed {the sick woman/her} answered the ringing phone himself in the hospital room since she was so sick.

MANAGE

39. The company president was glad that the sensible woman who supervises {the industrious man/him} wrote the glowing recommendation herself for the prestigious position that the man had applied for.

OVERCHARGE

40. The accounting office noticed that the underhanded man who overcharged {the likable woman/her} wrote the critical invoice himself on the company letterhead late at night.

REFERENCES

Abney, Steven. 1987. *The English Noun Phrase in its sentential aspects*. Unpublished PhD dissertation, MIT. Distributed by MIT Working Papers in Linguistics.

Abney, Steven. 1989. A Computational Model of Human Parsing. *Journal of Psycholinguistic Research* 18: 129-144.

Abney, Steven, and Mark Johnson. 1991. Memory requirements and local ambiguities for parsing strategies. *Journal of Psycholinguistic Research* 20, no. 3: 233-250.

Aho, Alfred, and Jeffrey Ullman. 1972. *The Theory of Parsing, Translation, and Compiling*. Prentice-Hall.

Alphonce, Carl, and Henry Davis. 1997. Motivating Non-directional Movement. In *Rightward Movement,* ed. van Riemsdijk, LeBlanc, and Beermann (John Benjamins).

Altmann, G. 1988. Ambiguity, Parsing Strategies and Computational Models *Language and Cognitive Processes* 3: 73-98.

Altmann, G., and M. Steedman. 1988. Interaction with Context during Human Sentence Processing. *Cognition* 30: 191-238.

Babyonyshev, Maria, and Edward Gibson. 1995. Processing Overload in Japanese. *MIT Working Papers in Linguistics* 26: 1-35.

Badeker, William, and Kathleen Straub. 1999. The processing role of structural constraints on the interpretation of pronouns and anaphors. Johns Hopkins University/Villa Julie College manuscript.

Bader, Markus. 1994. Syntactic-function ambiguities. *Folia Linguistica* 28, no. 1-2: 5-66.

Bader, Markus, and Inge Lasser. 1994 German verb-final clauses and sentence processing: Evidence for immediate attachment. In *Perspectives on sentence processing,* ed. Clifton, Frazier, and Rayner (Lawrence Erlbaum).

Bever, Thomas. 1970. The cognitive basis for linguistic structures. In *The making of cognitive science: essays in honor of George A. Miller*, ed. Hayes (Cambridge University Press).

Boland, Julie, Michael Tanenhaus, and Susan Garnsey. 1989. Evidence for the Immediate Use of Verb Control Information in Sentence Processing. *Journal of Memory and Language* 29, no. 4: 413-432.

Browning, Margaret. 1987. *Null operator constructions*. PhD dissertation, MIT. Distributed by MIT Working Papers in Linguistics.

Chomsky, Noam. 1970. Remarks on Nominalization. In *English Transformational Grammar,* ed. Jacobs and Rosenbaum.

Chomsky, Noam. 1973. Conditions on Transformations In *A Festschrift for Morris Halle,* ed. Anderson and Kiparsky (Holt, Rinehart and Winston).

Chomsky, Noam. 1981. *Lectures on Government and Binding: The Pisa Lectures*. Foris.

Chomsky, Noam. 1986. *Barriers*. MIT Press.

Chomsky, Noam. 1995a. Bare Phrase Structure. In *Government and Binding Theory and the Minimalist Program,* ed. Gert Webelhuth (Blackwell).

Chomsky, Noam. 1995b. *The minimalist program*. MIT Press.

Clark, H. H., and C. J. Sengul. 1979. In search of referents for nouns and pronouns. *Memory and Cognition* 7: 35-41.

Clifton, Charles, and Lyn Frazier. 1989. Comprehending Sentences with Long-Distance Dependencies. In *Linguistic Structure in Language Processing*, ed. Carlson and Tanenhaus (Kluwer).

Cornell, Thomas. 1995. On the relation between Representational and Processing Models of Asyntactic Comprehension. *Brain and Language* 50: 304-324.

Crain, Stephen, and Janet D. Fodor. 1985. How Can Grammars Help Parsers? In *Natural Language Parsing: Psychological, Computational and Theoretical Perspectives*, ed. Dowty, Kartuunen, and Zwicky (Cambridge University Press).

Crain, Stephen, and Mark Steedman. 1985. On not being led up the garden path: the use of context by the psychological parser. In *Natural Language Processing: Psychological, Computational and Theoretical Perspectives*, ed. Dowty, Karttunen and Zwicky (Cambridge University Press).

Crocker, Matthew. 1996. *Computational Psycholinguistics: An Interdisciplinary Approach to the Study of Language*. Kluwer.

Crocker, Matthew. Mechanisms for Sentence Processing. In *Language Processing*, ed. Garrod and Pickering (UCL/MIT Press), to appear.

Daneman, Meredyth, and Patricia Carpenter. 1980. Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior* 19, no. 4: 450-466.

Evers, Arnold. 1986. Clause Union in French and German. *Groninger Arbeiten zur Germanistischen Linguistik* 28: 170-201.

Ferreira, Fernanda, and Charles Clifton. 1986. The independence of syntactic processing. *Journal of Memory and Language* 25: 348-368.

Ferreira, Fernanda and J.M. Henderson. 1990. The use of verb information in syntactic parsing: A comparison of evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16: 555-568.

Fischer, Beth, and Murray Glanzer. 1986. Short-term storage and the processing of cohesion during reading. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology* 38, no. 3-A: 431-460.

Fodor, Janet D., and Lyn Frazier. 1980. Is the human sentence parsing mechanism an ATN? *Cognition* 8, no. 4: 417-459.

Fodor, Janet D., and Atsu Inoue. 1994. The Diagnosis and Cure of Garden Paths. *Journal of Pscyholinguistic Research* 23, no. 5: 407-434.

Fodor, Janet D., and Atsu Inoue. 1998. Attach Anyway. In *Reanalysis in Sentence Processing,* ed. Fodor and Ferreira (Kluwer).

Frank, Robert. 1992. *Syntactic Locality and Tree-adjoining Grammar: Acquisition and Processing Perspectives*. Unpublished PhD dissertation, University of Pennsylvania.

Frazier, Lyn. 1978. *On Comprehending Sentences: Syntactic Parsing Strategies*. Unpublished PhD dissertation, University of Connecticut.

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

Frazier, Lyn. 1990. Identifying Structure Under X^0 . In *Yearbook of Morphology 3*, ed. A. Jongman and A. Lahiri.

Frazier, Lyn, and Charles Clifton. 1998. Sentence Reanalysis and Visibility. In *Reanalysis in Sentence Processing*, ed. Fodor and Ferreira (Kluwer).

Frazier, Lyn, Charles Clifton and J. Randall. 1983. Filling gaps: Decision principles and structure in sentence comprehension. *Cognition* 13: 187-222.

Frazier, Lyn, and Janet D. Fodor. 1978. The sausage machine: a new two-stage model of the parser. *Cognition* 6: 291-325.

Frazier, Lyn, and Keith Rayner. 1982. Making and correcting errors during sentence comprehension: eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology* 14: 178-210.

Gibson, Edward. 1991. *A Computational Theory of Human Linguistic Processing: Memory Limitations and Processing Breakdown.* Unpublished PhD dissertation, Carnegie-Mellon University.

Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68: 1-76.

Gibson, Edward, Neal Pearlmutter, Enriqueta Canseco-Gonzalez, and Gregory Hickok. 1996. Recency Preference in the Human Sentence Processing Mechanism. *Cognition* 59: 23-59.

Gorrell, Paul. 1987. *Studies in Human Syntactic Processing: Ranked-parallel versus Serial Models.* Unpublished PhD dissertation, University of Connecticut.

Gorrell, Paul. 1995. *Syntax and parsing*. Cambridge University Press.

Grimshaw, Jane (1986). "Subjacency and the S/S' Parameter."*Linguistic Inquiry*, 17, no. 2: 364-369.

Haarmann, Henk, Marcel Just, Patricia Carpenter. 1997. Aphasic Sentence Comprehension as a Resource Deficit: A Computational Approach. *Brain and Language* 59, no. 1: 76-120.

Haarmann, Henk, and Herman Kolk. 1991. A Computer model of the temporal course of Agrammatic Sentence Understanding: The Effects of Variation in Severity and Sentence Complexity. *Cognitive Science* 15: 49-87.

Huang, James. 1982. *Logical Relations in Chinese and the Theory of Grammar*. PhD Dissertation, MIT. Distributed by MIT Working Papers in Linguistics.

Inoue, Atsu, and Janet Dean Fodor. 1995. Information-paced Parsing of Japanese. In *Japanese Sentence Processing,* ed. Reiko Mazuka and Noriko Nagai (Lawrence Erlbaum).

Jackendoff, Ray. 1977. *X' Syntax: A Study of Phrase Structure*. MIT Press.

Johnson-Laird, Phillip. 1983. *Mental Models: towards a cognitive science of language,inference, and consciousness.* Harvard University Press.

Just, Marcel, and Patricia Carpenter. 1992. A capacity theory of comprehension: individual difference in working memory. *Psychological Review* 99, no. 1: 122-149.

Just, Marcel, Patricia Carpenter, and Jacqueline Woolley. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 111, no. 2: 228-238.

Kayne, Richard. 1983. Connectedness. *Linguistic Inquiry* 14: 223-250.

Kimball, John. 1973. Seven Principles of Surface Structure Parsing in Natural Language. *Cognition* 2, no. 1: 15-47.

King, J. and M. A. Just. 1991. Individual differences is syntactic processing: the role of working memory. *The Journal of Memory and Language* 30: 580-602.

Konieczny, Lars. 1996. *Human sentence processing: a semantics-oriented parsing approach*. Unpublished PhD dissertation, Universit t Freiburg.

Konieczny Lars, Barbara Hemforth, Christoph Scheepers, and Gerhard Strube. 1997. The role of lexical heads in parsing: Evidence from German. *Language and Cognitive Processes* 12, no. 2-3: 307-348.

Konieczny, Lars, Barbara Hemforth, and Gerhard Strube. 1991. Psychologisch fundierte Prinzipien der Satzverarbeitung jenseits von Minimal Attachment. *Kognitionswissenchaft* 2: 58-70.

Kurtzman, H. 1985. *Studies in Syntactic Ambiguity Resolution*. Unpublished PhD dissertation, MIT.

Larson, Richard. 1988. On the double object construction. *Linguistic Inquiry* 19, no. 3: 335-391.

Lasnik, Howard and Mamuro Saito. 1984. On the nature of proper government. *Linguistic Inquiry* 15: 235-289.

Lewis, Richard. 1993. *An Architecturally-based Theory of Human Sentence Comprehension*. PhD dissertation, Carnegie Mellon University. Available as CMU Technical Report #CMU-CS-93-226.

Lyons, John. 1968. *Introduction to Theoretical Linguistics*. Cambridge University Press.

MacDonald, Maryellen, Marcel Adam Just, and Patricia Carpenter. 1992. Working Memory Constraints on the Processing of Syntactic Ambiguity. *Cognitive Psychology* 24: 56-98.

MacDonald, Maryellen, Neal Pearlmutter, and Mark Seidenberg. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101: 676-703.

Manzini, M. Rita. 1992. *Locality: a theory and some of its empirical consequences*. MIT Press.

Marcus, Mitch. 1980. *A Theory of Syntactic Recognition for Natural Language*. MIT Press.

Marcus, Mitch, Donald Hindle, and Margaret Fleck. 1983. D-Theory: Talking about Talking about Trees. *Proceedings of the 21st Meeting of the ACL*.

Marslen-Wilson, William. 1973. Linguistic structure and speech shadowing at very short latencies. *Nature* 244: 522-523.

Marslen-Wilson, William. 1975. Sentence perception as an interactive parallel process. *Science* 189: 226-228.

Mazuka, Reiko, and Kenji Itoh. 1995. Can Japanese Speakers Be Led Down the Garden Path? In *Japanese Sentence Processing,* ed. Mazuka and Nagai (Lawrence Erlbaum).

Melvold, Janis. 1991. Factivity and Definiteness. In *MIT Working Papers in Linguistics* 15.

Milward, David. 1994. Dynamic Dependency Grammar. *Linguistics and Philosophy* 17: 561-605.

Cuetos, F. and D.C. Mitchell. 1988. Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition* 30, no. 1, 73-105.

Mitchell, D.C and V. Holmes. 1985. The role of specific information about the verb in parsing sentences with local structural ambiguity. *Journal of Memory and Language* 24, 542-559.

Nicol, Janet, and David Swinney. 1989. The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research* 18, 5-19.

Pareschi, Remo, and Mark Steedman. 1987. A lazy way to chart-parse with categorial grammars. *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*.

Partee, Barbara, Alice ter Meulen and Robert Wall. 1993. *Mathematical Methods in Linguistics.* Kluwer.

Pearlmutter, Neal, and Maryellen MacDonald. 1995. Individual Differences and Probabilistic Constraints in Syntactic Ambiguity Resolution. *Journal of Memory and Language* 34: 521-542.

Pearlmutter, Neal, and Aurora Mendelsohn. 1999. Serial versus parallel sentence comprehension. Northeastern University manuscript.

Phillips, Colin. 1996. *Order and Structure*. PhD Dissertation, MIT. Distributed by MIT Working Papers in Linguistics.

Phillips, Colin. Linear Order and Constituency. *Linguistic Inquiry*, to appear.

Phillips, Colin, and Edward Gibson. 1997a. Local Attachment and Competing Constraints. Poster presented at 10th Annual CUNY Conference on Human Sentence Processing, Santa Monica, CA. Available at <http://www.ling.udel.edu/colin/research/ftp.html>.

Phillips, Colin, and Edward Gibson. 1997b. On the Strength of the Local Attachment Preference. *Journal of Psycholinguistic Research* 26, no. 3: 323-346.

Pollard, Carl, and Ivan Sag. 1994. *Head-driven Phrase Structure Grammar*. University of Chicago Press and CSLI Publications.

Pritchett, Bradley L. 1987. *Garden Path Phenomena and the Grammatical Basis of Language Processing.* PhD dissertation, Harvard University.

Pritchett, Bradley L. 1988. Garden Path Phenomena and the Grammatical Basis of Language Processing. *Language* 64, no. 3: 539-576.

Pritchett, Bradley L. 1991. Head Position and Parsing Ambiguity. *Journal of Psycholinguistic Research* 20: 251-270.

Pritchett, Bradley L. 1992. *Grammatical Competence and Parsing Performance*. University of Chicago Press.

Rayner, Keith and Lyn Frazier. 1987. Parsing temporarily ambiguous complements. *Quarterly Journal of Experimental Psychology* 39A: 657-673.

Resnik, Phillip. 1992. Left-Corner Parsing and Psychological Plausibility. *Proceedings of the Fourteenth International Conference on Computational Linguistics (COLING '92).*

Richards, Norvin. 1997. *What moves where in which language?* PhD Dissertation, MIT. Distributed by MIT Working Papers in Linguistics.

Richards, Norvin. 1998. The Principle of Minimal Compliance. *Linguistic Inquiry* 29, no. 4: 599-629.

Rizzi, Luigi. 1990. *Relativized Minimality*. MIT Press.

Roark, Brian, and Mark Johnson. (1999). Efficient probabilistic top-down and left-corner parsing. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics*.

Rosenkrantz, S.J. and P.M. Lewis II. (1970). Deterministic left corner parsing. In *IEEE conference record of the 11th annual symposium on switching and automata.*

Ross, J.R. 1967. *Constraints on variables in syntax*. PhD Dissertation, MIT.

Rudin, Catherine. 1988. On Multiple Questions and Multiple Wh Fronting. *Natural Language and Linguistic Theory* 6: 445-501.

Schneider, David, and Colin Phillips. 1999. *Reanalysis as a Last Resort?* CUNY Conference on Human Sentence Processing. New York, NY.

Sch tze, Carson. 1995. "PP Attachment and Argumenthood. *MIT Working Papers in Linguistics* 26.

Sch tze, Carson, and Edward Gibson. 1999. Argumenthood and English Prepositional Phrase Attachment. *Journal of Memory and Language* 40, no. 3, 409-431

Stabler, Edward. 1993. *The logical approach to syntax: foundations, specifications, and implementations of theories of government and binding.* MIT Press.

Stabler, Edward. 1994. The finite connectivity of linguistic structure. In *Perspectives on Sentence Processing*, ed. Clifton, Frazier, and Rayner (Lawrence Erlbaum).

Steedman, Mark. 1989. Grammar, interpretation and processing from the lexicon. In *Lexical Representation and Process*, ed. Marslen-Wilson (MIT Press).

Steedman, Mark. 1993. Categorial Grammar. *Lingua* 80: 221-258.

Steedman, Mark. 1996. *Surface Structure and Interpretation*. MIT Press.

Steedman, Mark. *The Syntactic Process*. University of Edinburgh manuscript, to appear from MIT Press.

Stevenson, Suzanne. 1994. *A Competitive Attachment Model for Resolving Syntactic Ambiguities in Natural Language Parsing*. Unpublished PhD dissertation, University of Maryland.

Stowe, Laurie. 1986. Parsing WH-constructions: evidence for on-line gap location. *Language and Cognitive Processes* 1, no. 3: 227-245.

Stowell, Tim. 1981. *Origins of phrase structure*. Unpublished PhD dissertation, MIT.

Straub, Kathleen, and William Badecker. 1999. A priming study of structural accessibility and co-reference processing. CUNY Conference on Human Sentence Processing, New York, NY.

Sturt, Patrick. 1997. *Syntactic Reanalysis in Human Language Processing*. Unpublished PhD dissertation, University of Edinburgh.

Sturt, Patrick, and Matthew Crocker. 1996. Monotonic syntactic processing: A cross-linguistic study of attachment and reanalysis. *Language and Cognitive Processes* 11, no. 5: 449-494.

Sturt, Patrick, and Matthew Crocker. 1999. Generalized Monotonicity for Reanalysis Models. In *Reanalysis in Sentence Processing* , ed. Fodor and Ferreira (Kluwer).

Sturt, Patrick, Martin Pickering, and Matthew Crocker (1999a). Exploring the 'Reanalysis as a Last Resort' Strategy. CUNY Conference on Human Sentence Processing, New York, NY.

Sturt, Patrick, Martin Pickering, and Matthew Crocker (1999b). Structural Change and Reanalysis Difficulty in Language Comprehension. *Journal of Memory and Language* 40, 136-150.

Szabolsci, Anna, and Frans Zwarts (1991). *Unbounded dependencies and Algebraic Semantics*. Handout from the Third European Summer School in Language, Logic and Information, niversit t des Saarlandes, Saarbr cken, Germany.

Tabor, Whitney, Cornell Juliano, and Michael Tanenhaus. 1997. Parsing in a Dynamical System: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. Language and Cognitive Processes, to appear.

Tabor, Whitney, and Michael Tanenhaus. Dynamical Models of Sentence Processing. *Cognitive Science*, in press.

Tanenhaus, Michael, and Michael Spivey-Knowlton. 1996. Eye-tracking. *Language and Cognitive Processes*11, no. 6: 583-588.

Tanenhaus Michael, Michael Spivey-Knowlton, Kathleen Eberhard, and Julie Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268 (June 16), 1632-1634.

Trueswell, John, and Michael Tanenhaus. 1991. Tense, temporal context and syntactic ambiguity resolution. *Language and Cognitive Processes* 6: 303-338.

Trueswell, John, Michael Tanenhaus, and Christopher Kello. 1993. Verb-specific constraints in sentence processing: separating lexical effects from garden paths. *Journal of Experimental Psychology: Learning, Memory and Cognition* 19, no. 3: 528-553.

Wanner, Eric. 1980. The ATN and the Sausage Machine: Which one is baloney? *Cognition* 8, no. 2: 209-225.

Warner, J., and A. L. Glass. 1987. Context and distance-to-disambiguation effects in ambiguity resolution: Evidence from grammaticality judgments of garden path sentences. *Journal of Memory and Language* 26: 714-738.

Weinberg, Amy. 1993. Parameters in the theory of sentence processing: Minimal commitment theory goes east. *Journal of Psycholinguistic Research* 22, no. 3: 339-364.

PARSING AND INCREMENTALITY

by

David Andrew Schneider

An abstract of a dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy with a major in Linguistics.

Summer 1999

Approved: ____

Colin Phillips, Ph.D.
Professor in charge of dissertation

ABSTRACT

There is a great deal of evidence that language comprehension occurs very rapidly. To account for this, it is widely, but not universally, assumed in the psycholinguistic literature that every word of a sentence is integrated into a syntactic representation of the sentence as soon as the word is encountered. This means that it is not possible to wait for subsequent words to provide information to guide a word s initial attachment into syntactic structure. In this dissertation I show how syntactic structures can be built on a word-by-word incremental basis.

A psycholinguistically plausible theory of parsing should generalize to all languages. In this work I show how both head-initial and head-final languages can be parsed incrementally. There is a significant amount of temporary ambiguity in head-final languages related to the fact that heads of constituents are not available until the end of the phrase. This temporary ambiguity hinders incremental parsing in many frameworks. Underspecification of the features of a head allows for incremental structuring of the input in head-final structures, while still retaining the temporary ambiguity that is so common in these languages. Featural underspecification is extended to categorial features; I do not assume that every head must always be specified for its category.

I assume that the incremental parser builds structures in accord with the principles of the grammar. In other words, there should be no need to submit a structure built by the parser to a separate grammar module to determine whether or not the sentence obeys the grammar. As one aspect of this, I show how wh-movement phenomena can be accommodated within the theory. As part of the treatment of wh-movement, constraints on wh-movement are incorporated into the system, thereby allowing the difference between grammatical and ungrammatical wh-movement to be captured in the parse tree.

In addition to being incremental and cross-linguistically generalizable, a parsing theory should account for the rest of human parsing behavior. I show that a number of

the structurally-motivated parsing heuristics can be accommodated within the general parsing theory presented here. As part of the investigation of the incremental parser, experimental evidence is presented that establishes a preference for structure-preserving operations in the face of temporary ambiguity. In particular, the experiments show that once a commitment has been made to a particular analysis of a verbal argument, there is a preference to avoid reanalyzing the argument. This preference holds even though the reanalysis is not particularly difficult, and the analysis that is adopted in preference to the reanalysis disobeys a general parsing preference for attachments to recent material. Thus, it appears that existing structural assumptions are rejected only as a last resort.