Syntactic Reanalysis in Human Language Processing

Patrick Sturt

PhD
University of Edinburgh
1997

Declaration

I declare that this thesis has been composed by myself and that the research reported here has been conducted by myself unless otherwise indicated.

Patrick Sturt
Edinburgh, May 13, 1997

View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Edinburgh Research Archive

Acknowledgements

I would firstly like to thank my supervisors Matt Crocker and Martin Pickering, for all their advice and support, without which my research work would have required a great deal of needless reanalysis. I would also like to thank all the other people with whom I have had helpful and interesting discussions about the work presented here over the last three and a half years. These include, but are not limited to Holly Branigan, Chuck Clifton, Janet Fodor, Paul Gorrell, Barbara Hemforth, Masayo Iida, Yuki Kamide, Lars Konieczny, Simon Liversedge, Vincenzo Lombardo, Don Mitchell, Wayne Murray, Peter Sells, and the members of the Sentence Processing workshop in Edinburgh and Glasgow.

I extend special thanks to my examiners, Bob Frank and Martin Corley, whose intellectual midwifery resulted in a safe delivery of this thesis.

Some of the research reported here has been, or will soon be, published elsewhere. We will indicate this at the relevant points in the text.

The questionnaire studies reported in chapter 8 were conducted in collaboration with Holly Branigan and Yoko Matsumoto Sturt. I would like to thank Kuabrat Branigan and the Park Language School in Sheffield for invaluable help in running Experiment 1, and Jim Hutton and the Institute of Applied Language Studies in Edinburgh for similar help in Experiment 2. I would also like to thank Steffan Corley for help in running the questionnaire reported as Experiment 2 in chapter 7.

I would also like to thank my office mates, Saturnino Luz and Siamak Rezaei for helping to create a surreal working environment.

My colleagues at Wentworth have been a constant source of inspiration for the research reported in this thesis, and I extend special thanks to them all, in particular Lizzie Birdsworth, Noeline Bourke, Rita Connors, Erica Davidson, Jim Fletcher, Joan Ferguson, Greg Miller and Bea Smith.

Finally, I would like to thank my wife Yoko for being able to live with an absent-minded PhD student, and for reminding me that there is more to life than psycholinguistics.

Abstract

This thesis combines theoretical, computational and experimental techniques in the study of *reanalysis* in human sentence comprehension.

We begin by surveying the main claims of existing theories of reanalysis, and identify *representation preservation* as a key concept. We show that the models which most obviously feature representation preservation are those which have been formulated within the *monotonicity framework*, which assumes that there are aspects of representation which are updated monotonically (i.e. non-destructively) from state to state, and that any reanalysis which requires a non-monotonic update is predicted to cause processing disruption.

Next, we present a computational implementation, based on the monotonic theory of Gorrell (1995b). We argue that in constructing such a model of reanalysis, it is essential to consider not only declarative constraints, but also the computational processes through which reanalysis is performed. In contrast to Gorrell (1995b), we provide a computational implementation which makes reanalysis routines explicit, leading to novel predictions in cases where there exist more than one alternative for structural revision. I show why preferences for such reanalysis ambiguities may differ between predominantly head initial languages such as English, and head final languages such as Japanese.

After this, we consider the empirical consequences of the implemented model, in particular in relation to recent experimental data concerning modifier attachment. We show that the model is too restrictive, and we argue that the appropriate way to expand its coverage is to apply the monotonicity constraints not directly to *phrase structure*, but to *thematic structure*. We provide a general framework which allows such non-phrase structural models to be defined, maintaining the same notion of monotonicity that was employed in the previous model. We go on to provide solutions to some computational problems which accompany this change.

Finally, we present two experimental studies. The first of these considers the issue of *reanalysis ambiguity*, and specifically the existence of a recency preference in reanalysis, as predicted by our implemented model. The predicted recency preference is confirmed in off-line tasks, such as comprehension accuracy and a questionnaire experiment, but is not confirmed in self-paced reading. We discuss some possible reasons for this dissociation between the on-line and off-line results. The second experimental study considers the effect of modifier attachment in Japanese relative clause ambiguities. In this study, we confirm the influence of thematic structure on the resolution of Japanese relative clause ambiguities, and we argue that this effect should be interpreted in terms of a constraint on reanalysis.

Contents

Chapter 1

Introduction

Natural language is vastly ambiguous; the number of possible readings for a sentence can increase exponentially with the length of the input. Given this fact, the human ability to recover the meaning of utterances efficiently and apparently effortlessly, must be seen as remarkable. Much psycholinguistic research has focussed on the decisions people make when initially faced with an ambiguous portion of the input. However, given the vastness of ambiguity in natural language, it cannot be guaranteed that the initially preferred analysis will always be consistent with the intended overall interpretation of the utterance. This means that a full theory of syntactic processing needs to specify not only how people initially resolve ambiguities, but also how they recover (or *reanalyse*) when the initially chosen analysis turns out to be incorrect. Reanalysis is a fundamentally complex process; at any point where reanalysis is possible, there may be any number of previous choice points, each of which may be revisable in any number of ways. Given this complexity, it is unlikely that reanalysis processes are random or arbitrary. Rather, it is more likely that these processes are guided and constrained by systematic principles and preferences. It is the goal of a theory of reanalysis to specify what these principles and preferences are, and this thesis may be seen as a contribution towards this goal.

1.1 Overview of the thesis

In chapter 2, we establish the background of reanalysis within psycholinguistic research, identifying key concepts and discussing existing proposals in the light of these. We begin by defining what we mean by the term reanalysis, and establish the need for a theory of reanalysis in psycholinguistic research, pointing out that reanalysis has hitherto played only a secondary role in all but a small number of experimental studies. We then move on to a discussion of the nature of reanalysis, identifying two of its key components, *diagnosis* and *cure*. We show how current models can be classified in terms of these two notions. We go on to discuss the concept of *representation preservation*; that is, the preference for the processor to maintain aspects of the structure that it has already built at the point where reanalysis becomes necessary. Current models are shown to differ in the extent to which representation preservation is assumed to be a constraint on reanalysis processes. We end the chapter with a more detailed discussion of two contrasting models.

In chapter 3 the focus moves to the class of *monotonic models*, which represent a particularly clear, and therefore theoretically interesting version of the representation preserving hypothesis. We show how the notion of monotonicity yields a precise definition of "preservation" in representation preserving models, and we discuss models which have been proposed within this framework, concentrating particularly on that given in Gorrell (1995b). We point out the necessity for monotonic theories such as Gorrell's to consider the actual processes involved in parsing and reanalysis, rather than simply defining the declarative constraints of the monotonic theory in question.

Chapter 4 describes a computational implementation of a monotonic model which takes the constraints of Gorrell (1995b) as its starting point. We show how the required constraints can be captured by assuming three basic parsing operations, one of which, *tree-lowering*, is used for monotonic reanalysis. We point out the existence of *reanalysis ambiguities*, in which the tree lowering operation can be applied at more than one site, and argue that theories of reanalysis must include preferences for such constructions. On the basis of examples from English and Japanese, we argue that such preferences may differ between predominantly head-initial languages such as English, and predominantly head-final languages such as Japanese, and suggest some reasons why this may be so.

Chapter 5 considers some of the empirical consequences of the model described in the previous chapter, particularly in the light of recent experimental data concerning modifier attachment. We show that the model is too restrictive, and argue that the appropriate way to expand its coverage is to apply the monotonicity constraints not directly to *phrase structure*,

as in classical monotonic theories, but to *thematic structure*. We give two general definitions of monotonicity, which allow one to maintain the notion of monotonic preservation while altering the types of representation to which it applies. Our *thematic monotonicity* constraint is the result of applying these general definitions to certain aspects of thematic structure. We show how the *thematic monotonicity* constraint combines insights from both Pritchett (1988) and Frazier and Clifton (1996).

In chapter 6, we consider computational problems arising from the implementation of the thematic monotonicity model. Unlike in chapter 4, where the goal was to produce an implementation for a *specific* model, the focus here is on *generality*. Specifically, we provide a general method for fully incremental parsing, and consider the problem of implementing general monotonicity definitions, as constraints on reanalysis.

In chapter 7, we describe a self-paced reading experiment and a questionnaire study designed to test a preference for the resolution of reanalysis ambiguity. Specifically, we test constructions in which reanalysis is necessary, but where there are two possible ways to reanalyse. In terms of the model described in chapter 4, this corresponds to a choice between applying *tree-lowering* either at a low (recent) or high (non-recent) site, and we predicted a low preference for these constructions. This low preference is confirmed in two off-line measures (comprehension accuracy and questionnaire responses), but not in reaction times in a self-paced reading experiment. We conclude the chapter by discussing some reasons for the distinction between the on-line and the off-line results.

In chapter 8, we describe two questionnaire experiments designed to test the influence of thematic structure on the resolution of Japanese relative clause ambiguities. Previous studies have found a high attachment preference for these constructions (Kamide and Mitchell, 1996). We argue that, in such left-branching structures, the preferred high attachment is derived via reanalysis. In our questionnaire studies we find that thematic domain boundaries reduce the number of high attachments, as predicted by the thematic monotonicity model described in chapter 5.

1.2 Caveats

The preceding section gives an idea of what the reader can expect to find in this thesis. In this section, we will give some indications about what the reader will not find here.

First, the reader will notice that, in common with much of the literature on psycholinguistic research, the syntactic structures we assume are rather naïve. For example, we employ neither the complex feature structures of Head Driven Phrase Structure Grammar (Pollard and Sag, 1994) nor the highly baroque structures of functional projections and moved elements typical of contemporary Chomskyan syntax (Chomsky, 1992). Instead, we have concentrated primarily on *processing* phenomena, and limited our *grammatical* assumptions to a fairly basic and general level.[1] We believe that the resulting account could be recast in terms of a number of syntactic theories. For example, it would be possible to build the reanalysis theory described here into a Principles and Parameters parser, provided that the grammar used by this parser is monostratal and non-derivational, as is the case in the parser described in Crocker (1996). A similar account in Head Driven Phrase Structure Grammar may also be possible, along the lines of the model described in Konieczny (1996). The monotonicity constraints which we develop here rely to a great extent on the *constituency* based notions of dominance and precedence. However, we believe that correlates of these constraints could also be defined in terms of *dependency* structures, allowing the possibility of building a version of the theory based on a dependency theory, such as Word Grammar (Hudson, 1994; Hudson, 1990).[2] Specializing the parser in terms of one or other of the available syntactic theories would bring challenges of its own, and would doubtless lead to novel and interesting predictions. However, we feel that it is possible, and often desirable, to investigate processing phenomena from a standpoint that is reasonably neutral with respect to syntactic theory, just as it usually considered desirable to investigate grammatical phenomena from a standpoint that is neutral with respect to processing theory.

As far as the processing model itself is concerned, we will be discussing *syntactic* effects on reanalysis. It is widely assumed that reanalysis processes make use of a variety of knowledge sources, including semantics, pragmatics, statistical preferences and intonation. In concentrating on chiefly the *syntactic* aspects of reanalysis, our model will necessarily be incomplete,

[1]We will clarify points where the choice of syntactic analysis has particularly important consequences.

[2]Building a fully incremental parser for *Minimalist* theory (Chomsky, 1992) would be a considerable undertaking, since the notion of derivation on which standard minimalism crucially depends is fundamentally non-incremental, and does not correspond to what we know of processing in either comprehension or production. This would appear to necessitate a fairly indirect grammar-parser relation, which would limit the theoretical interest of the enterprize. Much more feasible and interesting from the processing point of view is the dynamic version of minimalist theory proposed by Phillips (1995; 1996), where the derivations are performed from left-to-right, and are intended to correspond precisely to the stages of incremental parsing.

and therefore may miss some important generalizations. However, the task of building a working model which incorporates all of the knowledge sources that may influence reanalysis is probably AI complete, since it would require the model to draw inferences from world knowledge. It seems a more reasonable goal to concentrate initially on a small well-defined subpart of the problem, and subsequently to investigate how it interacts with other components of the process.

Furthermore, the incremental nature of the model we present here can, we believe, act as a basis for the investigation of other, non-syntactic effects on reanalysis. For example, it is known that the frequency of certain constructions can affect preferences in attachment ambiguities and in reanalysis. In order to define such statistical preferences in a psychologically interesting way, it is desirable to have a realistic model of the contexts in which they apply during real human parsing, and this necessitates an algorithm which builds connected structure incrementally. Incrementality is also a prerequisite for modelling the effects of semantic interpretation, since it is well known that plausibility effects can make themselves known very quickly (Traxler and Pickering, 1996). However, determining exactly how semantics and pragmatics influence the reanalysis process is an interesting and complex question which we leave to future research.

Chapter 2

Experimental and Theoretical Research on Reanalysis

2.1 Analysis and Reanalysis

Psycholinguistics is an interdisciplinary field of enquiry bringing together techniques of experimental psychology, linguistics and computer science to study the ability of humans to acquire, comprehend and produce language, and the problems involved in language breakdown.[1] The field involves the study of linguistic processing at a number of levels, from the very fine grained details of word recognition to the higher level processes of discourse interpretation. A central part of psycholinguistics, which is referred to as "sentence processing" or "parsing", involves the study of how people compute hierarchical structure and dependency relations between words and phrases, in order to derive a compositional semantic interpretation of an utterance. An important assumption, which has been borne out by a large volume of research, is that these combinatorial processes are guided and constrained by grammatical knowledge. It is therefore assumed that, at some level, there is a component of sentence comprehension which can be viewed as the assignment of a grammatical structure, or an *analysis*, to a string of words. For example, given a sentence such as *Jack thinks Bill is silly*, the grammatical analysis might include among other things, the information that *thinks* is the main verb of the sentence, that *Jack* is its subject, and that *Bill is silly* is its clausal complement. Given

[1]Much of the material in sections 2.2 and 2.5 of this chapter will appear in Sturt and Crocker (to appear).

the extent of ambiguity in natural language, it is clear that a string of words may sometimes be compatible with more than one analysis, as in the sentence *Jack said Bill died yesterday*, which is compatible both with analysis in which *yesterday* modifies *died*, and with an analysis in which *yesterday* modifies *said*. This kind of ambiguity is known as *global ambiguity*. Another type of ambiguity, which will play a central part in this thesis is *local ambiguity*. A string is regarded as locally ambiguous if, given some processing algorithm,[2] more than one analysis is compatible with the input at some intermediate stage of processing, regardless of whether or not all of these analyses will remain compatible with the input as a whole. So, for example, in the string, *Jack knows Bill is silly* (making certain assumptions about the processing algorithm), at the point where *Bill* is read, the input is compatible both with an analysis in which *knows* takes a noun phrase direct object (a slot which can be occupied by *Bill*), and also with an analysis in which *knows* takes a clausal complement (the subject slot of which can be occupied by *Bill*). Because only the second of these analyses is compatible with the input as a whole, the sentence is regarded as *locally*, but not *globally* ambiguous. The sentence *Jack knows Bill very well* exhibits the same local ambiguity, but is globally compatible only with the NP direct object analysis.

A *serial* model of sentence processing (such as that in Frazier and Rayner (1982), which we will discuss in section 2.3) applies preference strategies at ambiguous points in the input, to select just one analysis. Clearly, if just one analysis is selected when more than one alternative is possible, then there will be occasions in which this analysis is incompatible with the intended interpretation, or, in some cases, incompatible with the global grammatical structure of the utterance. In such cases, the processor must "recover" the correct analysis in order to arrive at the intended interpretation.

In general, the process of discarding the current analysis and recovering the correct analysis is known as *reanalysis*. As we shall see in this introductory chapter, although a theory of reanalysis is necessary for a complete model of parsing, the study of reanalysis is still very much in its infancy. Experimental psycholinguistics has concentrated on the decisions that people initially make to resolve ambiguities, and reanalysis has played an important role in

[2]As Abney and Johnson (1991) point out, local ambiguity is not an intrinsic property of strings, but a property of strings *given* some processing algorithm. When we use the term "locally ambiguous" in this thesis, we will do so on the assumption of a fully connected incremental parsing algorithm, as appears to be the general practice in psycholinguistics.

the methodology which has been used. However, very few experimental studies have looked at reanalysis as a research topic in its own right. Meanwhile, on the theoretical side, models have been proposed that attempt to explain intuitive differences in the difficulty of reanalysis, but these have typically not considered in any great detail the computational processes by which reanalysis is carried out.

However, before we go on to discuss the specific models that have been proposed, we will begin by offering a definition of reanalysis.

2.2 A definition of "Reanalysis"

As we said above, "reanalysis" can be loosely defined as the discarding of one analysis in favour of another. However, in order to be precise, we need a definition which will allow us to choose when one analysis can be counted as being different from another. Different models make different implicit assumptions on this question, and therefore, what counts as "reanalysis" varies considerably from model to model. For example, in Gibson's (1991) restricted parallel system, "reanalysis" refers only to recovery from "parsing breakdown" situations, which are predicted to occur in those cases where a previously discarded parse has to be re-introduced. The re-ordering of alternatives still under parallel consideration, which presumably must occur at some level, is not regarded as reanalysis. Similarly, in "constraint satisfaction" models (e.g. MacDonald, Pearlmutter, and Seidenberg (1994)), which are characterized by the parallel competitive activation of multiple alternatives, minor changes in relative activation levels are not usually classed as "reanalysis", but some researchers working in this tradition (Trueswell, Tanenhaus, and Kello (1993)) do not deny the existence of "true garden path" effects, which are indeed associated with "reanalysis". Presumably, reanalysis in a constraint satisfaction model corresponds to the re-activation of an alternative whose activation has dropped below some threshold to a level close to zero.

By contrast, within a serial framework, the *Garden Path* model (Frazier (1978), Frazier and Rayner (1982)) assumes that normal processing is accompanied by any number of minor reanalyses, as initial attachments based on restricted grammatical knowledge are discarded by higher level processing modules. However, even within the Garden Path tradition, it is not always clear which varieties of structural change count as reanalysis. For example, in models influenced by the Garden Path tradition a distinction is often made between the

attachment of postmodifiers of *verbs*, which are incorporated into the representation via *sister adjunction*, and the attachment of postmodifiers of *nouns*, which are incorporated via *Chomsky adjunction* (see Clifton, Speer, and Abney (1991) for an example of this distinction). Adams (1995) uses this distinction to argue that the attachment of NP modifiers is, in some sense, more destructive than attachment of VP modifiers, and therefore, in this model, we should presumably see attachment of NP modifiers as involving some form of reanalysis, and attachment of VP modifiers as not.

Given this wealth of subtly different nuances associated with the term, we propose to clarify what *we* mean by "reanalysis", before discussing the notion any further.

We view the processing of a string as a series of *states*, where each word induces a transition from one state to the next. The construction of a grammatical analysis can be viewed as a process of *dependency* formation, where *dependencies* can be viewed as directed binary relations, usually between heads and the items which the modify or select. Dependencies take different forms in different grammatical theories. For example in categorial grammars (Lambek, 1958; Moortgat, 1988), dependencies are represented by argument cancellation. Meanwhile, Head Driven Phrase Structure Grammars, (Pollard and Sag, 1994), supplement argument cancellation with feature structure unification. In Government and Binding theory (Chomsky (1981; 1986a)) dependencies take the form of *licensing* relations (e.g. case marking, theta role assignment), and more recent versions of Chomskyan syntactic theories (Chomsky, 1992), employ feature "checking". In Word Grammar (Hudson (1994; 1990)), dependencies between words are theoretical primitives. Despite the varieties of formal devices which are used to model the notion, we believe that dependency is fundamental to all syntactic theories, and plays an important part in human language comprehension. We will employ this notion in our definition of reanalysis.

We will use the abbreviation Dep^S to stand for the following proposition: "Dep is the set of dependencies holding in the preferred analysis at state S". We will abbreviate the notion of "a transition from state S to state T" as $S \Longrightarrow T$. Now, we define reanalysis as follows:[3]

A transition $S \Longrightarrow T$ involves *reanalysis* iff $Dep^S \nsubseteq Dep^T$.

From the definition, it can be seen that by "reanalysis" we mean "breaking at least one

[3] Note that the definition makes the simplifying assumption that it will always be possible to identify one particular transition where the preference for one alternative is replaced by the preference for another.

dependency". From this it follows that, for example, attaching a postmodifier is not a case of reanalysis, since it involves *adding* rather than *breaking* a dependency. On the other hand, in a ranked parallel system, changing the relative preference order of analyses in such a way that the preferred analysis at S is no longer the preferred analysis at T *will* count as reanalysis, if there are any dependencies at S which do not hold at T.

2.3 The Importance of Reanalysis in Psycholinguistics

The ultimate aim of psycholinguistics must be to gain a full understanding of the process by which sound waves (or sequences of written characters) are mapped to and from representations of meaning. So far, progress in this endeavour has been limited to relatively small parts of these processes. In sentence processing research, the majority of experimental studies have been directed at the investigation of the initial decisions that people make to resolve local ambiguities. Although, as we will see below, reanalysis plays a central part in the methodology used to investigate this issue, very few experimental studies have looked seriously at reanalysis as a phenomenon in its own right.

The dominant methodology[4] for studying initial ambiguity resolution strategies can be traced back to the influential eye-tracking experiments reported in Frazier and Rayner (1982), which we will briefly describe in this section. The main goal of Frazier and Rayner's study was to demonstrate the existence of serial strategies in human sentence processing, and to argue in particular for the strategies of *Minimal Attachment* and *Late Closure*. We will take the strategy of Late Closure as an example here, considering some of the authors' original materials. Consider (2.1) below:

(2.1) a. Since Jay always jogs a mile this seems like a short distance to him.

b. Since Jay always jogs a mile seems like a short distance to him.

Late Closure predicts that the processor prefers to attach incoming material to recently built structure. In (2.1), late closure predicts that the NP *a mile* is preferentially attached as a direct object of the preceding verb *jogs*, rather than as a subject of the main clause. Assuming a serial control strategy, in which the processor locally commits itself to only one

[4]By "methodology", we mean the general technique of eliciting garden paths, rather than eye-tracking in particular.

resolution of the ambiguity and subsequently reanalyses if necessary, then it is predicted that any information which disconfirms the initially chosen Late Closure attachment will trigger reanalysis, which will result in increased processing difficulty. In the case of (2.1), this set of hypotheses predicts that there will be a difference in processing difficulty between (2.1.a) and (2.1.b) in the region beginning with *seems*. Specifically, greater processing difficulty will be detected in (2.1.b), where the initial Late Closure attachment is disconfirmed, than in (2.1.a), where it is confirmed. In fact, the authors found evidence for both Late Closure and Minimal Attachment, by demonstrating statistically significant increases in reading time, and in the number of eye regressions in the predicted dispreferred structures.

Clearly, the methodology described in Frazier and Rayner's study relies crucially on finding reanalysis effects in order to demonstrate ambiguity resolution strategies. However, as we will see below, Frazier and Rayner also considered another side to the question, namely that of *how* the processor reanalyses. Although Frazier and Rayner's study, and the methodology it exemplifies, has been hugely influential, subsequent experimental research has almost exclusively followed up issues related to the first question of what influences initial ambiguity resolution, rather than the second question of how the processor reanalyses. Of course, the question of what influences initial ambiguity resolution is an important one, because its answer promises to shed light on a number of fundamental properties of mental architecture. One important question is the nature of the knowledge informing the process of ambiguity resolution. For example, are ambiguities initially resolved on the basis of all possible relevant information, including semantic/pragmatic plausibility, fine grained statistics and syntactic well-formedness (c.f. MacDonald, Pearlmutter, and Seidenberg (1994), Trueswell, Tanenhaus, and Kello (1993)), or does the processor initially make its decisions based on restricted low-level knowledge sources, such as structural preferences, and subsequently refine or revise these decisions when higher level information becomes available (c.f. Frazier and Rayner (1982) Rayner, Carlson, and Frazier (1983), Ferreira and Henderson (1990), Mitchell, Corley, and Garnham (1992)).

However, in order to gain a fuller understanding of human language processing, we need to supplement our knowledge of initial ambiguity resolution with knowledge about reanalysis. The language processor often initially resolves ambiguity in a way which is inconsistent with the final interpretation of the utterance, which means that initial resolutions of attachment ambiguities cannot be totally reliable predictors of the global reading of an utterance. This point can be seen more clearly if we consider certain dissociations between results from on-line ambiguity resolution experiments and results from corpus studies. One example of such a dissociation is reported in Brysbaert and Mitchell (1996), and involves Dutch relative clause attachment ambiguities of a form similar to *Someone shot the servant of the actress who was on the balcony*. Although the author's eye-tracking experiment showed clear evidence for a commitment at the earliest measurable point to the high attachment of the relative clause (i.e. the reading where the servant is on the balcony), a search of a Dutch corpus revealed a strong low attachment bias for structures of this type (i.e. sentences where the intended meaning corresponds to the reading where the actress was on the balcony predominated). If the on-line evidence reflects a preference for initial ambiguity resolution, and the corpus used by the authors reflects a realistic sample of the language to which Dutch speakers are exposed, then, in the majority of cases, reanalysis will be necessary in order to derive the intended interpretation of these constructions. Although the experiment was designed to test a separate issue (the effect of linguistic experience on initial ambiguity resolution) results such as these also illustrate that initial ambiguity resolution is unlikely to be a reliable predictor of the final interpretation of an utterance. A full understanding of language comprehension must include a consideration of how alternatives are recovered when an initial decision is discarded. In short, we need a theory of *reanalysis*, which would make predictions regarding the the relative accessibility of alternatives to the initial decision, and how the processor chooses when more than one alternative is available. This is necessary so long as the processor ever adopts the wrong initial analysis, which is likely to be a frequent occurrence in sentence comprehension.

2.4 The Nature of Reanalysis

Reanalysis is an inherently complex process as compared with initial ambiguity resolution. In order to see this, consider Figure 2.1. The point we are going to make is easiest to see if we restrict our attention to a maximally incremental processing system, where each word is incorporated into single totally connected syntactic representation before the processor moves on to consider the next word. As we will argue in later chapters of this thesis, such a notion of incrementality is a desirable property for models of the human syntactic processor,

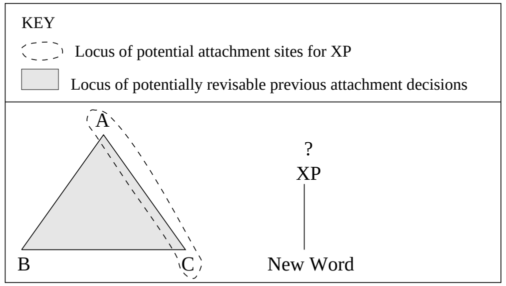


Figure 2.1:

and is assumed, either implicitly or explicitly, in many theories. Consider the problem of incorporating a new word into the representation, as informally illustrated in figure 2.1. Assuming that the processor attaches in accordance with the theory of trees, so that it cannot create crossing branches, or allow a node to have more than one mother,[5] then the locus of initial *attachment* possibilities for the projection of the new word corresponds with the right frontier of the tree built so far—that is, depending on the processing system used, the constituent XP could be attached as a daughter of any of the nodes on the path from C to A, or alternatively, the root A could be attached as the left daughter of XP (see Stevenson (1994a), (1994b) and the model given in chapter 4 for particularly clear examples of "right-frontier based"[6] attachment strategies). However, no other attachments are possible if the processor wants to obey the theory of trees, retain a connected structure, and preserve all dependencies.

Now consider the possibilities for *reanalysis*. It may be the case that XP cannot find a felicitous attachment site on the right frontier. If this is true, then, according to most models, the processor will attempt to revise a previous decision. In contrast to the situation of initial ambiguity resolution, there are no *a priori* structural constraints on either the locus of revisable previous choice points or on the locus of possible alternatives for any one of these

[5]The theory of trees is introduced on page 48, chapter 3. See also Wall (1972) and Partee, ter Meulen, and Wall (1993) for further discussion.

[6]The *right frontier* of a tree can be defined as the set of all nodes in the tree which do not precede any other node.

choice points. Given the inherent complexity of reanalysis, it is implausible to suggest that reanalysis is an essentially arbitrary or random process. Instead, it is highly likely that the process is guided by systematic constraints and strategies. A definition of these strategies is essential to a full model of human parsing, since even if we successfully identify the preferences that humans exhibit in initially resolving ambiguities, this information is incomplete without some further consideration of how these initial decisions are revised. Frazier and Rayner's (1982) study, which we mentioned above, made some initial progress towards the study of strategies in reanalysis. Specifically, they considered the following three hypotheses about the reanalysis mechanism:

Backward Reanalysis: The processor proceeds backwards, attempting alternative resolutions of each choice point, in reverse chronological order (note that this chronological backtracking algorithm is used by the PROLOG theorem prover, and is therefore the default algorithm used by parsers written in this language).

Forward Reanalysis: The processor abandons its parse of the sentence, and returns to the beginning of the sentence to start again (presumably the parser would have to retain at least some details of the original parse, to avoid making the same mistake twice).

Selective Reanalysis: The processor makes use of the information it has available in order to prioritize its search for alternative solutions. For example, in (2.1b), the disambiguating word *seems* lacks a subject. The processor can make use of this kind of information in order to start looking for an NP which could act as the missing subject. This kind of information can be used to find the most relevant choice points to look at during reanalysis.

Frazier and Rayner argued that the observed patterns of eye movements supported the third of these alternatives, the *selective reanalysis hypothesis*. Given the extent of ambiguity in natural language, and the temporal and computational constraints on processing, we would expect the search for previous choice points to be prioritized in a way which is likely to maximize the chances of finding the correct solution. A simple chronological search would not, in the general case, guarantee this. Therefore, some version of the *selective reanalysis* hypothesis is likely to be true of the human language processing system. However, selective reanalysis is a high level description of a strategy which could be instantiated in any number of different

ways. The question of exactly *how* the processor selects its choice point for reanalysis is an important one, which has only recently begun to be considered in any great depth.

A further question which must be answered in a theory of reanalysis is that of what makes some kinds of reanalysis difficult. If the parser applies systematic strategies in its reanalysis processes, then it is likely that there are systematic and predictable differences in the accessibility of different types of reanalysis. Hence, any patterns of difficulty which we observe can be used to infer information about the nature of the constraints which apply to reanalysis. For example, in a recent series of experiments, Ferreira and Henderson (1991) found that reanalysis difficulty is significantly increased when the head of an ambiguous phrase is separated from its disambiguating region, so that (2.2a) is more difficult to reanalyse than (2.2b):

(2.2) a. While the boy scratched the dog that is hairy yawned loudly.

b. While the boy scratched the big and hairy dog yawned loudly.

Ferreira and Henderson interpreted this finding in terms of a processor which assigns thematic roles in a head driven manner, so that the PATIENT role of *scratched* is assigned to the NP headed by *dog* earlier in (2.2a) than in (2.2b), allowing time for the activation of the globally correct subcategorization frame to decay in (2.2a), increasing the difficulty of the reanalysis. Although this interpretation of the data has been disputed (see Hemforth, Konieczny, and Scheepers (1994a)), the study demonstrates that reanalysis difficulty is amenable to experimental investigation, and suggests a fruitful avenue of future research, which has not been followed up to any great extent.

2.5 Theoretical Models of Reanalysis

In this section, we will turn our attention to the abstract sub-components of the reanalysis process, and we will consider in more detail how theoretical models have proposed constraints on these processes. Specifically, we will identify two dimensions on which theories of reanalysis may vary, and classify existing models according to them. We will argue that many of the models of reanalysis that have been proposed to date are incomplete in that they concentrate on abstract specifications of constraints on reanalysis, without paying due attention to the actual computational processes involved. The theoretical approach to reanalysis has its origins in computational and theoretical linguistics. Most of the models that we discuss in this section have as their goal the formulation of constraints on reanalysis, such that certain types of reanalysis are predicted to be systematically more difficult than others.

2.5.1 The Components of the Reanalysis Process

Consider example (2.3):

(2.3) The wedding guests saw the cake was still being decorated.

Assume that *the cake* is initially attached as the direct object of *saw*, and that the processor attempts to reanalyse this decision at *was*. Assume also that the processor is serial to the extent that it does not have the sentential complement reading available[7]. Then, the reanalysis must involve at least the following operations:

1. *the cake* must be "detached" from its current position as the direct object of *saw* (we will call this position the *detachment site*).
2. *the cake* must be re-attached as into the subject of *was* (we will call this position the *target site*).
3. The (projection of) the new word *was* must be attached into a position in the representation (we will call this position the *attachment site* of the new word).

As we stated above, recent models of reanalysis have concentrated on the aim of accounting for differential costs associated with different reanalyses by postulating various constraints on the possibilities for revision. The proposed constraints may vary along two dimensions; firstly, models vary in the extent to which they constrain the *diagnosis* versus the *cure* of garden paths (terms we have borrowed from Fodor and Inoue (1994)). Secondly, models vary in the nature of the computational limitations which are assumed to be the ultimate motivation for the constraints. Specifically, while some models attribute their constraints directly to limits in the computational resources available to the processor (such as bounded working memory, for example), others express the computational limitations less directly, by appealing to a

[7] In a parallel system, such as that of Gibson (1991), the processor simply has to discard the structure corresponding to the direct object reading from its working set, and attach *was* into the structure corresponding to the sentential complement reading.

preference encoded in the processor to *preserve* certain core aspects of representation. We will call the former class *resource bounded* and the latter class *representation preserving* models. It is with the problem of defining *representation preserving* models that this thesis will be concerned.

2.5.2 Diagnosis vs. Cure

As we mentioned above, there is still not a great deal of experimental evidence available on the nature of reanalysis, and many researchers have relied on intuitive distinctions, between degrees of difficulty associated with different types of reanalysis. An example of such a distinction, originating in the work of Pritchett (1988), is that between "conscious" garden paths (which, typically are claimed to cause noticeable processing difficulty) and "unconscious" garden paths (which, intuitively cause only a minor disruption to processing). Contrast (2.3), repeated below, which would on most accounts be classed as an *unconscious*, or easy garden path, with (2.4), which on many would be classed a *conscious*, or difficult garden path (c.f. Pritchett (1988), (1992), Gorrell (1995b)).

(2.3) The wedding guests saw the cake was still being decorated.

(2.4) While the wedding guests ate the cake was still being decorated.

Using the intuitive distinctions such as these as a clue, researchers have attempted to define constraints on reanalysis which exclude garden paths from which recovery is difficult, but include easy garden paths.

It should be noted before we proceed with the discussion that many researchers disagree with making a distinction between conscious and unconscious garden paths. Frequently voiced objections include the possibility that the difficulty of reanalysis may vary according to a continuous function as opposed to a binary distinction, as well as the fact that experimental techniques have not yet been found which can probe whether or not a reanalysis reaches consciousness. We do not propose to justify the conscious/unconscious distinction here. However, we do point out that while the theories that we will be discussing in this section predict that, all other things being equal, certain types of reanalysis are systematically easier to achieve than others, they do not necessarily commit themselves to defining this distinction in terms of *consciousness*, despite the fact that many researchers in the field have done so previously.

The prediction that certain types of reanalysis are systematically easier than others leads to a number of further predictions which allow experimental testing—predicted easy reanalyses should be preferred to predicted difficult reanalyses in situations where both are possible, all other things being equal; comprehension accuracy and reading speed should be adversely affected in sentences which demand a reanalysis which is predicted to be difficult relative to sentences which allow a reanalysis which is easy, for example.

However, it does not make sense to conduct such experiments without a model to test, and we believe that it is reasonable to base such initial models on intuitions. Throughout this thesis, we will freely cite the intuitions of researchers who have worked previously in the field, making the assumption that these intuitive data correspond to actual processing preferences which should be captured in a model. Although we do provide two experimental studies of reanalysis issues in chapters 7 and 8, there is a need for more extensive experimental investigation, which we leave for future research.

Recall from the previous section that reanalysis typically involves finding a *detachment site*, finding a *target site* to re-attach the detached constituent, and finding a *new attachment site* for (the projection of) the new word. The constraints in the following models can be defined in terms of one or more of these notions:

Pritchett (1988) The target site must be in the same *theta-domain* as the detachment site.[8]

Pritchett (1992) The detachment site must govern or dominate the target site.

Gorrell (1995b) The detachment site must dominate the target site. (n.b. this is a consequence of the indelibility of dominance and precedence relations, rather than being stipulated).

Lewis (1993) The detachment site must be *local* to the new attachment site of the incoming word (where the notion of *local* corresponds to "within the same maximal projection").

Fodor and Inoue (to appear) The detachment site must be accessible to the new attachment site of the incoming word through a chain of grammatical dependencies.

[8]For a definition of *theta-domain*, see page 2.6.1.

Stevenson (1994a), (1994b) Similar constraints to those proposed in chapter 4, but derived from the architecture of a hybrid Competitive Activation model.

It can be seen from the above that the models differ with respect to the relative importance attached to the *detachment site* and the *target site*. For example, Lewis (1993) and Fodor and Inoue (to appear) impose no constraints on the *target* site, beyond those that apply to initial attachment, but impose strong constraints on the *detachment site* with respect to the position where the projection of the new word is attached. In terms of search, this implies that the search for the attachment error that is the ultimate cause of the need for reanalysis is more tightly constrained than the search for an alternative attachment. On the other hand, Pritchett (1988), (1992) imposes no constraints on the position of the detachment site itself, but does impose constraints on the position of the *target site* relative to the detachment site. Pritchett's model implies that the search for an alternative attachment is more tightly constrained than the search for the error. We can call models of the former kind, which constrain the search for the error more heavily than the search for alternative attachments, "diagnosis-constrained models", and models of the latter kind, which constrain the search for alternative attachments more heavily than the search for the error, "cure-constrained models". Of course, many models constrain elements of both diagnosis and cure; for example, the model which we will describe in chapter 4 imposes a right-frontier based accessibility constraint on the search for a *detachment site*, and to this extent is *diagnosis-constrained*. On the other hand, the model also strongly constrains the space of possible *target positions*, and to this extent it is also *cure*-constrained. This shows that diagnosis-constrained models and cure-constrained models are not mutually exclusive. Rather, diagnosis and cure should be seen as two aspects of search which, in any particular processing model, may independently be constrained to a greater or lesser extent. To the extent that a model can be seen to constrain diagnosis at the expense of cure, that model can be called diagnosis-constrained model, and vice versa.

2.5.3 Resource-Boundedness vs. Representation Preservation

It is likely that the ultimate motivation for the constraints on reanalysis is computational. Given the combinatorial explosion of ambiguity in natural language, and the *a priori* complexity of search involved in reanalysis which we discussed above, it is unlikely that all alternatives to an initially chosen analysis are equally accessible to the processor. However, as we have seen above, models vary with respect to how these limitations manifest themselves in processing behaviour.

Gibson (1991) is a clear example of a *resource-bounded* model, since in this model, the bounds on computational resources play a direct role in the constraints imposed on reanalysis. In Gibson's model, the processor maintains multiple analyses in parallel, but, to avoid memory overload, prunes away analyses which exceed a certain processing cost relative to the preferred analysis. The necessity for re-introducing a pruned structure is predicted to correspond to the conscious garden path effect in human processing. In this system, the constraint on reanalysis is clearly motivated by working memory limitations. Conserving working memory also plays a key part in the more serial model of Lewis (1993), where attachment sites waiting to be filled are pruned from the parser's memory under certain conditions.

By contrast, *Representation Preserving* models attribute to the processor an unwillingness to destroy representation, or propose that there is a difficulty involved in doing so. The ultimate motivation for this may very well be limitations in computational resources; after all, representations presumably require computational resources to build and modify. However, the link between bounds on computational resources and constraints on processing is much less direct in these models than in the *resource bounded* models. Instead, the claim is that there is something about destroying structure which is intrinsically dispreferred. The model described in Fodor and Inoue (to appear) contrasts strongly with a representation preserving model, since it claims that there is no difficulty involved in destroying structure *per se*, just so long as the limited inferential capabilities of the processor allow it to determine *which* structure needs to be destroyed.

An example of a principle which is clearly motivated by representation preservation is Frazier and Clifton's (to appear) *Minimal Revisions* Principle (Frazier, 1990a; Frazier, 1994):

> **Minimal Revisions:** Don't make an unnecessary revision. When revision is necessary, make the minimal revision consistent with the error signal, maintaining as much of the already assigned structure and interpretation as possible.

As an abstract statement about the nature of human sentence processing, the concept behind the Minimal Revisions principle makes a great deal of sense. Clearly, the opposite claim, say a "Maximal Revisions" principle would be computationally implausible; it would entail

a strategy in which the processor throws away hard-earned interpretations for no good reason. Furthermore, the small amount of experimental evidence which exists points towards representation preserving behaviour on the part of the processor. For example, Ferreira and Henderson's (1991) results, as described, can be interpreted in terms of a processor which finds it hard to reanalyse dependencies which have been semantically confirmed.

However, the statement of Minimal Revisions, as given above, could be interpreted in various ways, depending on precisely what is meant by "minimal revision", and "maintaining as much of the already assigned structure and interpretation as possible." The question is, what general metric would be appropriate to quantify destructiveness to representations? One possibility, which would be consistent with our dependency-based definition of reanalysis, is to count the number of dependencies which have to be broken by the revision in question. This metric was behind the explanation we suggested for differences, described in Sturt and Crocker (1996), and in chapter 4 of this thesis, between reanalysis preferences in head-initial versus head-final languages.

However, the simple *number* of dependencies which are broken is unlikely to provide a *complete* answer to the question—in many cases, both "hard" and "easy" garden paths involve breaking only one dependency. In the reputedly "easy" garden path in (2.3) repeated below, one dependency (between *saw* and *the cake* is broken, and the same is true of the "hard" garden path in (2.4), where the broken dependency is between *ate* and *the cake*.

(2.3) The wedding guests saw the cake was still being decorated.

(2.4) While the wedding guests ate the cake was still being decorated.

Furthermore, there is evidence that the preferred interpretation of certain relative clause attachment ambiguities in Italian (De Vincenzi and Job (1995)) and Japanese (Kamide and Mitchell (1996)) may be derived via the breaking of one dependency even in the absence of any syntactic or semantic error signal (see Sturt and Crocker (1997), and chapter 5 of this thesis for discussion).

This leads us to another side to this question, which is the consideration that there might be certain *aspects* of representation that are, in some sense, more sacred to the processor than others, and which it is therefore more reluctant to destroy. This has been the basic assumption behind many of the reanalysis models that have been recently proposed. Thus, as we shall see, Pritchett's (1988) *Theta Reanalysis Constraint* encodes a preference to preserve certain aspects of *thematic* structure, as defined by his "thematic domains". Meanwhile, Gorrell (1995b) and Weinberg (1993), (1995), which we will consider in chapter 3, and which we will develop in subsequent chapters of this thesis encode a preference to preserve (description based) *constituent* structure representations. On the other hand, Bader (1996) argues that *multiple* representation types (a mixture of syntactic, lexical-morphological and prosodic) should be preserved.

As with the diagnosis/cure distinction, resource-boundedness and representation-preservation are not mutually exclusive characteristics. That is, for example, a processor may be resource bounded in some way, say, due to limitations on working memory, but may use representation-preservation as a heuristic for keeping the computational cost of reanalysis within these bounds. Similarly, a processor may be primarily motivated to preserve representations, but be further limited by rather modest inference capabilities.

2.6 A More Detailed Description of Two Models

In this section, we will look at two recent models, and examine in more detail how they constrain reanalysis. The first model we look at is that of Pritchett (1988, 1992), which exemplifies a representation preserving, cure based approach. Secondly, we look at the proposal of Fodor and Inoue (to appear), which exemplifies a diagnosis based approach.

We will postpone our discussion of monotonic models such as Gorrell (1995b) until the next chapter.

2.6.1 Pritchett (1988)

The Theta Reanalysis Constraint (TRC)

The motivation behind Pritchett's (1988)[9] work is to produce a parsing model whose behaviour is motivated, to as great an extent as possible, by the principles of the competence grammar. The competence grammar that Pritchett assumes is similar to the Principles and

[9]The TRC was replaced by the "On Line Locality Constraint" (OLLC) in subsequent work (Pritchett (1992)). However, we discuss the earlier version of the theory here, as it is a much clearer example of a representation preserving principle. In chapter 5, we will argue that, although the TRC has some empirical problems, these are not solved satisfactorily by the OLLC, and the TRC offers a better starting point for the definition of constraints on the reanalysis of modifier attachment.

Parameters model of grammar (Chomsky (1981; 1986b; 1986a)). The process of initial ambiguity resolution is assumed to be driven by the maximum satisfaction of licensing constraints, with particular emphasis being placed on the thematic component of the grammar. This means that the parser works in such a way as to minimize the number of thematic roles that are left unassigned or unreceived at any point in processing, a strategy which he calls *Theta Attachment*. Conversely, reanalysis is constrained by a principle which guarantees the preservation of the thematic structure that has been assigned. This constraint on reanalysis is called the *thematic reanalysis constraint*:

> **Theta Reanalysis Constraint:** (TRC) Syntactic reanalysis which interprets a θ-marked constituent as outside of its current θ-domain is costly.
>
> **θ-domain:** α is in the γ θ-domain of β iff α receives the γ θ-role from β or α is dominated by a constituent that receives the γ θ role from β.

We can see how the TRC predicts the difference in difficulty between (2.3) and (2.4). First, consider the easier (2.3):

(2.3) The wedding guests saw the cake was still being decorated.

The NP covering *the cake* is initially attached as the direct object of *saw*:

[....[$_{VP}$ saw [$_{NP}$ the cake]]

In this position, *the cake* receives the THEME role from *saw*. Subsequently, on the input of *was*, the THEME role is taken over by a clause (CP). Since *the cake* is dominated by this CP, it is still within the THEME theta domain, and the reanalysis is predicted to cause little difficulty.

[....[$_{VP}$ saw [$_{CP}$ [$_{IP}$ [$_{NP}$ the cake] was]]]]

Now consider (2.4), which is claimed to be a difficult reanalysis:

(2.4) While the wedding guests ate the cake was still being decorated.

We assume that *the cake* is assigned the THEME theta role by *ate*. However, the reanalysis required to reinterpret *the cake* as the subject of the matrix clause brings the constituent out of this domain (i.e. it no longer receives the *theme* role from *saw*, and is not dominated by a constituent which is assigned this role). Hence the reanalysis is predicted to cause difficulty.

Discussion

Pritchett's model can be clearly seen as a cure-based, representation preserving theory, in that the Theta Reanalysis Constraint can be interpreted as a preference to preserve the content of thematic domains. However, there are a number of aspects of the model which are unsatisfactory. Firstly, the theory makes no attempt to define the computational processes by which reanalysis is carried out. It is certainly reasonable to define declarative constraints on representational change, and it can be seen as a strength of Pritchett's theory that such constraints can be defined so simply. However, as we will argue in chapters 3 and 4, a full theory of reanalysis must say not only which types of reanalysis are possible, but also *how* reanalysis proceeds. Defining such dynamic processes leads to new empirical predictions which cannot be made simply with reference to static principles such as the Theta Reanalysis Constraint. Furthermore, the consequences of a reanalysis model are hard to assess without the aid of a computational implementation, given that reanalysis is an inherently complex phenomenon.

A second aspect of Pritchett's model which can be criticised is the fact that the licensing-based approach fails to make any interesting predictions regarding modifiers, which do not obviously fit into the scheme of licensing. Recent work originating in the well known study by Cuetos and Mitchell (1988) has made a number of interesting findings regarding modifiers, and these need to be accounted for in a theory of sentence processing. In fact, we will argue in chapter 5, that some of the accounts of these effects of modifier attachments, such as that of Frazier and Clifton (1996) can be reconciled with a version of the TRC. However, this requires a number of changes, and it is not obvious how the licensing-based approach would fit in with these.

A third aspect of Pritchett's model which has been the subject of criticism is his use of *head driven* parsing (Pritchett, 1991). This strategy, in which a constituent is not built until the head has been received in the input, is non-incremental, and is inconsistent with a wide range of evidence from head final languages (Hemforth, Konieczny, and Scheepers (1994a), Yamashita (1994), Bader and Lasser (1994)). However, since this aspect of Pritchett's model has been criticised elsewhere (Crocker (1996), Gorrell (1995b), Inoue and Fodor (1995)), we will not discuss the issue further in this chapter. We do, however, point out that head driven parsing is not a necessary consequence of the TRC; in fact, in chapter 5, we will describe

a model which combines certain key insights of the TRC with incremental, word by word structure building.

2.6.2 Fodor and Inoue

In contrast to Pritchett's model, the work of Fodor and Inoue (1994; to appear) represents a particularly clear example of a diagnosis-constrained approach to reanalysis. Indeed, the authors specifically claim that there is no difficulty involved in destroying structure *per se*, and that it is not the case that some cures are easier than others. They claim that difficulty only arises in cases where the processor's inference mechanism breaks down, after failing to find the source of the original parsing error that has led to the garden path. The value of Fodor and Inoue's model is that it allows us to test the consequences of assuming that representation preservation and constraints on cure are unnecessary in a theory of reanalysis. However, as we will see, although representation preservation and cure constraints are assigned a minimal role, there are points in the theory which rely crucially on the (often implicit) assumption of precisely such constraints.

In what follows, we will examine the proposal put forward in Fodor and Inoue (to appear) in some detail.

At the core of the reanalysis mechanism is the principle of *Attach Anyway*:

> **Attach Anyway:** Having established that there is no legitimate attachment site in the CPPM for the current input word, attach the input word into the CPPM wherever it least severely violates the grammar, and subject to the usual preference principles that govern *Attach*.[10]

The parser then tries to change the rest of the tree to legitimate the ungrammatical attachment produced by *Attach Anyway*. This process is governed by the principle of *Adjust*:

> **Adjust:** When a grammatical conflict has been created between two nodes or features X and Y in the CPPM, by either *Attach Anyway* or *Adjust*, eliminate the problem by altering minimally (i.e. no more than is necessary) whichever of X and Y was less recently acted on, without regard for grammatical conflicts thereby created between that node and other elements in the CPPM.

[10] *Attach* is simply the first pass attachment principle, which attempts to find a grammatical attachment site for the current input word.

It can be seen that *Adjust* is defined recursively. This results in a basic reanalysis algorithm that can be summarised as follows:

1. Attach the projection of the offending word into the tree via *Attach Anyway*.

2. Find X, a node or feature which is incompatible with the current configuration.

3. Change X in some way (e.g. if X is a node, reattach it elsewhere in the tree, if X is a feature, destructively change it).

4. If the resulting configuration is well-formed, then succeed, else go to 2.

Because of the recursive definition, steps 2 and 3 may be repeated any number of times, in a chain of inference that, in a successful reanalysis, will lead to the discovery of the original error source. Recalling our discussion earlier in the chapter, each application of step 2 may be seen as a *diagnosis* step in the inference chain, while each application of step 3 may be seen as a *cure* step in the inference chain. As we mentioned in the previous discussion, Fodor and Inoue's model can be seen as a strongly *diagnosis*-constrained model. What this means is that of steps 2 and 3, it is 3 to which the main constraint of the theory applies. The constraint in question is called the *Grammatical Dependency Principle*:

> **The Grammatical Dependency Principle (GDP):** When a grammar violation has been created in the CPPM by an action on node n in accord with *Attach Anyway* or *Adjust*, attempt to eliminate the problem by acting on a node that is grammatically incompatible with n.

The GDP ensures that each diagnosis step in the chain of inference is linked by a grammatical inconsistency. Given that the set of nodes or features which can be inconsistent with an attachment is highly constrained syntactically, this is equivalent to imposing a locality constraint, defined in terms of grammatical dependencies, on the diagnosis step of the search.[11] Let us see how a typical reanalysis would proceed in this framework. Consider the following "easy" sentence:

(2.3) The wedding guests saw the cake was still being decorated.

[11] Note that these local dependency links may be composed together to form a non-local dependency, as in long-distance extraction, for example.

In Fodor and Inoue's system, *the cake* is initially attached as the object of *saw*. When the auxiliary verb *was* is read, clausal structure is projected, and (presumably) the CP projection is attached as a sister of *saw*, via *attach anyway*, and as part of the attachment operation, incorporates the NP *the cake* as its subject, in a manner similar to the *tree-lowering* operation of Sturt and Crocker (1996), which we discuss in chapter 4[12]. The set of nodes with which the attachment of CP in this position is incompatible includes the verb *saw*, since it is labelled with a feature showing subcategorization for an NP. *Adjust* then changes this feature to one which subcategorizes for a CP. The lexicon is checked, and the repair is successful, since *saw* can subcategorize for a CP.

However, there is another aspect of Fodor the model which, in contrast to the Grammatical Dependency principle, is clearly a cure based, representation preserving principle. The authors call this the "Thematic Overlay Effect" (henceforth TOE). Informally, the content of TOE is that verbs and other thematic assigners resist having their arguments removed. This seems at first sight to be very similar to the TRC of Pritchett, but there are differences, the most important of which is that TOE does not apply in cases where the removal of an argument has been explicitly requested in an inference step; that is to say, the effects of TOE cannot be seen in cases where the steps of diagnosis inference follow the GDP. So, for example, in (2.3) above, since the GDP allows the processor to infer that it is necessary to remove *the cake* from *saw*, no TOE effects can be seen. However, this can be contrasted with the "difficult" garden path given in (2.5).

(2.5) While the boy scratched the big and hairy dog yawned loudly.

Here, there is no chain of grammatical inconsistencies leading from the lack of an overt subject for *yawned* to the misattachment of *the big and hairy dog* as the object of *scratched*.[13] This means that the processor has to apply an "emergency" reanalysis operation, which the authors call *theft*, and which applies when the GDP does not lead to a successful reanalysis. It is in cases of theft that TOE effects are seen. To make the point, Fodor and Inoue contrast (2.5) with a similar example, which involves *theft*, but does not induce the TOE effect:

(2.6) While the boy scratched the big and hairy dog and the cat yawned loudly.

[12] Thanks to Janet Fodor (personal communication), for giving some details of the attachment operations assumed in Fodor and Inoue (to appear).

[13] Note that this assumes a high "attach anyway" for *yawned*.

In (2.6), Fodor and Inoue argue that *yawned* prefers to take *the cat* as its subject, in preference to the entire coordinated NP. This leads to an analysis of (2.6) as an incomplete sentence, that could be continued as follows:

(2.7) While the boy scratched the big and hairy dog and the cat yawned loudly, the hamster continued running in its little wheel.

Fodor and Inoue point out that sentences such as (2.7) are intuitively easier to process than either (2.5) or the complete sentence analysis of (2.6), and that this is due to TOE. The reason for this is that *theft* removes the NP object in of *scratched* in (2.6), and forces the verb *scratched* to be re-interpreted as intransitive. In (2.7), in contrast, only one conjunt of the coordinated NP is removed, and although the internal structure of the coordinated phrase has to be revised, the verb *scratched* remains marked as a transitive verb, retaining an NP argument. Therefore no TOE effects are observed.

If the observation illustrated in (2.6) and (2.7) is correct, then it poses some interesting problems for most theories of reanalysis. However, it is instructive to note that in order to account for it, Fodor and Inoue appeal to a principle which is clearly representation-preserving and cure-constrained in its content. Indeed, it is difficult to see how such a contrast could be predicted on a purely diagnosis based account, as the diagnosis of a misattachment involving a coordinated NP would appear *a priori* to be *more* rather than *less* complex than one involving an uncoordinated NP.

Discussion

Fodor and Inoue's model is interesting and instructive in that it allows us to see the consequences of the extreme position they adopt. The model goes further than Pritchett's in that it considers the actual processes of inference in reanalysis. However, although it is possible to gain an intuitive idea about what kind of an algorithm is involved, there are still a number of issues which would have to be resolved before a computational implementation could be built to test the theory. For example, unless heavily constrained, the application of *Attach Anyway* could lead to serious intractability problems. Each "cure" step in a chain of adjustments may be executed as an ungrammatical attachment, and is not constrained by the GDP[14].

[14]Recall that the GDP applies to *diagnosis* rather than *cure steps.*

Furthermore, at the point when the processor makes the ungrammatical attachment, it may not know whether it is the *appropriate* ungrammatical attachment to make, because this may depend on the success of subsequent ungrammatical attachments. Given the fact that, for any reasonably sized phrase marker, the number of possible ungrammatical attachments at any one step in the chain of inference is quite large, the search space involved may be quite enormous.

2.7 Summary

In this chapter we have pointed out that, while a theory of reanalysis is fundamentally important for any model of human sentence processing, it has been the focus of surprisingly little research in psycholinguistics. In our survey of existing research on reanalysis, we have discussed, on the one hand, distinctions between cure based approaches and diagnosis based approaches, and on the other, distinctions between representation preserving and resource bounded models. We have argued that some version of the representation preserving hypothesis is very likely to be true of human language processing mechanisms. In the remainder of this thesis, we will look at the representation preserving hypothesis in considerable detail, looking at a class of *monotonic* models, which allow us to define exactly what is meant by "representation preservation", and to apply the concept to working processing models.

Chapter 3

The General Framework

3.1 The Monotonicity Framework

In the last chapter, we identified *representation preservation* as a property shared by a number of reanalysis models. However, the assumption of representation-preserving constraints on reanalysis leaves open the following two questions:

1. How should we formalize the notion of preservation?

2. Which aspects of representation should be preserved?

In this chapter,[1] we will see how the *monotonicity* framework provides an answer to the first question, and we will go on to consider the contribution to the monotonicity framework made by Gorrell (1995b). As far as the second question is concerned, we will see that the traditional approaches to monotonicity assume that it is *constituent* structure which is preferentially preserved. However, in chapter 5, we will define a method of generalizing monotonicity so that it can be applied to other representations.

What, precisely, do we mean by "representation preservation"? We can regard a representation as a set of (possibly binary) linguistic or structural relations. Now, recalling that a parse is viewed as a series of transitions from one state to the next, the notion of "preservation" implies that, once a relation has been added to the representation, it is not removed. *Monotonic* models (Gorrell (1995b), Weinberg (1993), Sturt and Crocker (1996)) use a form of

[1]Much of the material in section 3.1 of this chapter will appear in Sturt and Crocker (to appear) and Sturt and Crocker (1997). Much of the material in section 3.3 is drawn from Sturt and Crocker (1996).

representation called a *description* which can be "preserved" in this manner, while tolerating certain classes of *reanalysis*. Recalling our definition of reanalysis, this means that, while at the level of linguistic *dependencies*, a relation holding at some state S may fail to hold at the next state T, at the level of the *description*, all relations holding at state S also hold at state T.

3.2 Determinism and the Minimal Commitment Hypothesis

The monotonicity hypothesis has its origins in Marcus's (1980) attempt to apply to natural language processing the techniques of deterministic parsing, which had previously been used in the design of compilers for computer languages. Marcus saw determinism as a means of dealing with the pervasive problem of ambiguity in natural language, and argued that it provided an interesting alternative to the traditional techniques of parallelism and non-deterministic search. The idea of determinism as applied to natural language processing was that at ambiguous points in the input, the processor could "hedge its bets" until disambiguating information became available. In the case of Marcus's original work (Marcus (1980)), this was achieved by using a *look ahead* buffer to hold portions of the input, allowing the parser to wait for disambiguating information to become available.

It may be assumed that the human language processing system is subject to two opposing pressures. On the one hand, the processor is under pressure to build up a maximally informative interpretation word by word, and this, under standard assumptions, means that the parser is under pressure to make *commitments*. On the other hand, it may also be assumed that the processor is under pressure *not* to commit itself to decisions which may subsequently have to be revised, since such reanalysis presumably has a computational cost. One of the key tasks of a theoretical model of human language processing is to determine the parser's priorities when faced with these two competing pressures, and existing models can be characterised in terms of exactly where they stand on this issue. For example, the classical Garden Path model of Frazier (1978) and Frazier and Rayner (1982) may be characterised by the assumption of *maximal commitment*, as, at least at the constituent structure level, the parser always finds a single fully specified analysis at all points in the input, at the cost of routinely having to reanalyse its decisions.[2] On the other hand, as we will see below, many of the theories grew out of Marcus's work on deterministic parsing have been called *minimal commitment* models, because their use of delay mechanisms and underspecification is based on the (often implicit) assumption that, at least in many cases, the processor delays making commitments for the computational benefit of not having to reanalyse, even though making such commitments could otherwise allow a maximally informative incremental interpretation.

Clearly, an extreme version of the minimal commitment hypothesis would make for an implausible model of sentence processing, as it would imply that humans are never misled into making the wrong reading of a local ambiguity, and hence never have to reanalyse; a position which is contrary to a large body of experimental and intuitive evidence. Therefore, in order to be psychologically interesting, a minimal commitment model has to specify the *limits* to which the parser may hedge its bets. Marcus (1980) did this by limiting the size of the look-ahead buffer to three cells[3], thus forcing the parser to make a decision in cases where this limit was exceeded. However, given the considerable evidence for incremental interpretation, it is clear that this still results in a model which is too minimally committed to be useful as a psychological model (For examples of cases where Marcus's use of the delay mechanism leads to missed predictions, see Pritchett (1992)).

3.2.1 The Essence of Monotonicity

In later deterministic models, such as Marcus, Hindle, and Fleck (1983), Berwick and Weinberg (1984), Barton and Berwick (1985), the representations built by the parser were underspecified. So, as we will see below, for example, Marcus, Hindle, and Fleck (1983) used the transitivity of non-immediate dominance, as well as allowing node labels to be non-rigid designators.[4] In the case of Barton and Berwick (1985), underspecification involved the use of partial well-formed substring tables,[5] allowing the processor to delay its commitment concerning the syntactic category of a node covering some substring of the input. What these later deterministic models have in common is that the representation built by the parser is

[2]Note that later revisions of the Garden Path model, underspecification has been added for lexical category ambiguities (Frazier and Rayner, 1987) and modifier attachment (Frazier and Clifton, 1996).

[3]Roughly speaking, a buffer cell may either a single word, or a partial or complete noun phrase.

[4]This means that the description is underspecified as to whether or not two node labels refer to the same node.

[5]See Gazdar and Mellish (1989) for an introductory discussion of well-formed substring tables.

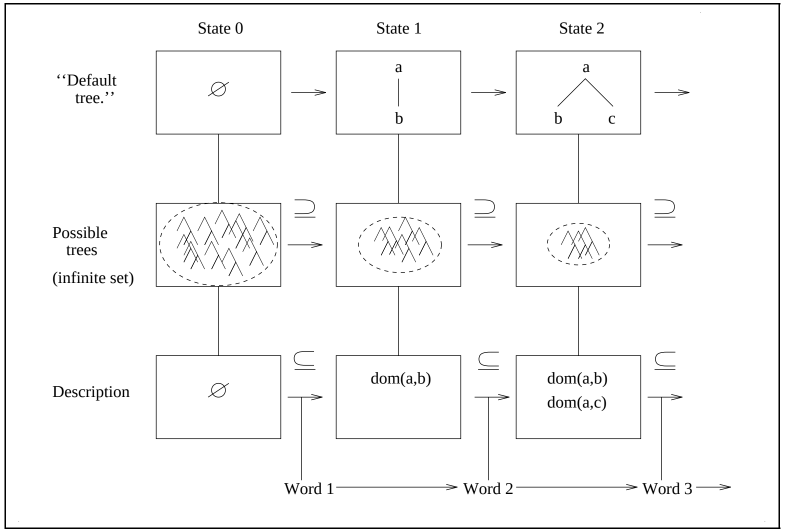


Figure 3.1: The Essence of Monotonicity.

monotonic; that is to say, new information can be *added* to the representation during the course of a parse, but it cannot be altered or removed. In what follows, we will discuss in some detail how these monotonic models work. We will see in the course of the discussion that, although the underspecified representations that are used are by definition *partial* structures, it is possible to reconcile them with a processing regime in which a *fully* specified representation is built up word by word, thus allowing for incremental interpretation to take place.

We can think of the dynamics of processing in monotonic models on three levels, as illustrated in figure 3.1. On the first level, the *description*, is a set of statements describing certain aspects of the syntactic representation. Following Marcus, Hindle, and Fleck (1983), the description standardly consists of a set of *dominance* relations; for example, at State 1, the description consists of the statement that a dominates b. As we shall see, the fact that the dominance relation is *transitive* (i.e. if X dominates Y and Y dominates Z, then X dominates Z) gives D-theory its flexibility and power. At any state, the description can be interpreted as an infinite set of "possible trees", which we have depicted on the middle level of the diagram, corresponding to all the trees which are consistent with the dominance statements, along with the axioms of the theory of trees. If we assume that the model does not employ *look ahead* buffers, stacks or other delay devices, then, at any point in processing, a minimal "default" tree can be generated from the description.[6] The ease with which such a default tree may be recovered from the description will depend on exactly how fully specified the parser's output is. In the case of the models which we will be discussing in this and the subsequent chapter, the "default" tree at some state S is simply the result of assuming all dominance relations in the description at S to encode *immediate* dominance, and assuming all nodes not mentioned in the description not to exist, that is, assuming the description is a "closed world". If we take State 1 in Figure 3.1, for example, the "possible trees" comprise the infinite set of trees corresponding to the statement that a dominates b , including trees in which other nodes exist on the domination path between a and b. However, at this state, the default tree contains only the unique member of the set of possible trees for which there is evidence in the description, namely, the tree in which a dominates b, and no other nodes exist; in other words, where a *directly* dominates b.

The constraint of monotonicity says that the description must be preserved—that is, once a dominance relation has been added, it cannot be deleted. This means that the description can be seen as a *monotonically increasing* set, where the description at each state will be a subset of all descriptions at subsequent states. Conversely, the "possible trees" comprise a *monotonically decreasing* set, in that adding new relations to the description removes "possible trees", but never adds them. Consider the effect of adding the new relation `dom(a,c)` at state 2. This has the effect of removing from the set of possible trees all structures in which `c` does not exist, for example, or all trees in which `c` dominates `a`. However, all the trees which are compatible with the description at state 2 are also compatible at state 1.

In this sense, the monotonicity framework shares common features with the classical *cohort* model of word recognition (Marslen-Wilson, 1987). The "cohort" in Marslen-Wilson's system is simply the set, at a given state of word recognition, of all word candidates consistent with the string of phonemes processed up to that point. As subsequent phonemes are recognized, word candidates which become inconsistent with the input are removed. If it subsequently

[6] This is what Marcus, Hindle, and Fleck (1983) called the "standard referent".

becomes necessary to reintroduce a candidate that has been eliminated at a previous state, then a processing cost is predicted. Thus, the cohort is a monotonically decreasing set, and is equivalent to the level of "possible trees" in the monotonicity framework. Of course, one of the major differences between syntax and word recognition is that, unlike the set of word candidates in a cohort, the set of continuations of a *syntactic* prefix is infinite, and thus impossible to enumerate.[7] The *description* in a monotonic processing model may therefore be viewed as an underspecified representation which, at any state S, encodes the locus of possible changes which the representation may undergo as new material is processed subsequent to state S.

3.2.2 Infinite Local Ambiguity

An illustration of the fact that a description can encode an infinite number of structures can be gained by considering the phenomenon of *infinite local ambiguity*. Infinite local ambiguity may arise in circumstances such as the following. Imagine the following string of words has been processed:

(3.1) I saw the man ...

Now, let us say that the minimal "default" interpretation contains the following information:

(3.2) (**State S:**) I $[_{VP}$ saw $[_{NP_x}$ the man$]]$

Because of the transitivity of the dominance relation, the set of "Possible Trees" will include trees in which various numbers of nodes intervene on the domination path between VP and NP_x. However, the minimal default tree will include only the single tree indicated above, in which VP *immediately* dominates NP_x. Now imagine that the input proceeds so that at state T, the default tree includes the following information:

(3.3) (**State T:**) I $[_{VP}$ saw $[_{NP_y}$ $[_{NP_x}$ the man$]$'s wife$]]$

By the time of state T, the default tree at state S has been removed from the set of Possible Trees, and, since the dependency between the verb *saw* and NP_x has also been removed, we class this change as a *reanalysis*. However, the new minimal interpretation at state T

[7] This distinction does not hold in languages which exhibit cyclical agglutinative morphology, where the number of word candidates for a prefix is potentially infinite.

has existed in the set of Possible Trees at all previous states, including that of State S. In fact, the input could continue indefinitely in the same manner, and the set of Possible Trees would at each stage be monotonically *narrowed down*, with the description simultaneously monotonically increasing.

(3.4) I saw the man's wife's friend's sister's ....

The type of local ambiguity exemplified in (3.1)-(3.4) is more pervasive than might be thought. We can expect to find similar examples in all cases where right branching and left-branching structures meet—a situation which, as pointed out by Inoue and Fodor (1995), is extremely pervasive in head-final languages such as Japanese. We will discuss the parsing of such constructions in more detail in chapters 4 and 6.

It should be clear at this point that the D-theory formalism is extremely powerful, in that it allows a compact encoding of an infinite number of possible trees. This property makes the monotonicity paradigm an interesting alternative to parallelism. In contrast to, for example, ranked parallel models (e.g. Gibson (1991)), in which alternative analyses are explicitly built, and structures pruned when they exhibit a particular relative or absolute processing cost, in the monotonic framework, multiple analyses are simply implicit in the description, and the "pruning" is simply the by-product of adding new relations to the description.

3.3 Gorrell and Structural Determinism

The model proposed by Gorrell (1995b) is important because it takes a position which is nearer to *maximal commitment*, and further from *minimal commitment* than any previous monotonic or deterministic model. This is due to the following properties which hold of Gorrell's model:

1. All sources of underspecification are removed, other than those which directly follow from the D-theory representational system. For example, unlike earlier models, such as Berwick and Weinberg (1984), Barton and Berwick (1985) and Weinberg (1993), Gorrell does not allow the parser to underspecify node category labels.

2. The *look ahead* buffers of previous deterministic models are removed, and incremental attachment behaviour is introduced in the parser.

3. An extra constraint is added to the D-theory system; specifically, the parser now has to preserve not only *dominance* relations, but also *precedence* relations, resulting in a more constrained model.

The particularly constrained nature of Gorrell's model is interesting for a number of reasons, the most important of which being the following:

1. As we will see, the lack of a look-ahead buffer, and the reduction of underspecification mean that it is now possible to reconcile monotonic models with the psychologically desirable property of *incrementality*.

2. As all other sources of delay and underspecification are removed, a theory of reanalysis difficulty can now be derived *directly* from the D-theory representational system.

It is crucial to note a further sense in which Gorrell's model differs from previous monotonic models, and that is that the requirement of monotonicity is applied purely to what Gorrell calls *primary relations* (i.e. dominance and precedence). By contrast, *secondary relations* (such as theta and case assignment, for example), are not constrained in this way, so that it is possible for linguistic relations to change over the course of a parse, while the dominance and precedence relations remain monotonically updated. This property means that, in contrast to earlier *minimal commitment* models, the D-theory conditions are being used not as a mechanism of underspecification that allows the parser to "hedge its bets", but as a way of limiting the accessibility of alternatives to the current preferred analysis.

In the following sections, we will describe Gorrell's model in some detail.

3.3.1 Conditions on Trees

We will say that the set of relations describing the phrase marker being constructed is *coherent* iff it conforms to the following conditions for trees (adapted from Partee, ter Meulen, and Wall (1993)).

1. **Single Root Condition:**

   There is a single node, the *root node*, which dominates every node in the tree:

   $\exists x \forall y.\ \mathrm{dom}(x, y)$

2. **Exclusivity Condition:**

   No two nodes can stand in both a dominance and a precedence relation:

   $\forall x, y.\ \text{prec}(x, y) \vee \text{prec}(y, x) \leftrightarrow \neg\ \text{dom}(x, y) \wedge \neg\ \text{dom}(y, x)$

3. **Inheritance:**

   (a.k.a. the "non-tangling" condition). All nodes inherit the precedence properties of their ancestors:

   $\forall w, x, y, z.\ \text{prec}(x, y) \wedge \text{dom}(x, w) \wedge \text{dom}(y, z) \rightarrow \text{prec}(w, z)$

Dominance and precedence are both defined as transitive relations. In addition, it is usually assumed that dominance is reflexive (every node dominates itself) and precedence is irreflexive. That is to say that dominance defines a *weak* partial order and precedence defines a *strict* partial order.

3.3.2 Primary and Secondary Relations

As we have mentioned above, Gorrell divides syntactic representation into *primary* relations (dominance and precedence) and *secondary* relations (theta-role assignment, c-command, case-assignment, etc), of which only the primary relations are constrained by monotonicity.[8] This is what he calls "structural determinism".

In the following, we again use the abbreviation "$S \Longrightarrow T$" to stand for the transition from state S to state T.

Thus, Gorrell's *Structural Determinism* constraint can be defined as follows:[9]

Dominance Constraint:

For each transition $S \Longrightarrow T$, if X dominates Y at state S, then X dominates Y at state T.

Precedence Constraint:

For each transition $S \Longrightarrow T$, if X precedes Y at state S, then X precedes Y at state T.[10]

[8]Note the notion of 'primary' versus 'secondary' relations here should not be confused with 'primary' versus 'secondary' phrases in Construal Theory (Frazier and Clifton, 1996).

[9]Note that although this constraint is equivalent to Gorrell's, it is not stated in identical terms.

[10]To avoid cluttering these definitions, we avoid explicit mention of the universal quantifiers over the variables X and Y. We continue to make this omission in subsequent definitions, where we believe there is no danger of misinterpretation.

Consider (2.3) and (2.4), repeated from the previous chapter:

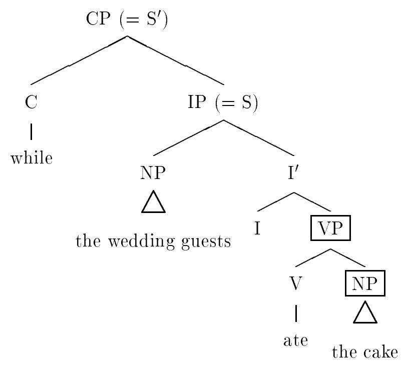
(2.3) The wedding guests saw the cake was still being decorated.

(2.4) While the wedding guests ate the cake was still being decorated.

Recall that (2.3) is standardly thought to be easier to process than (2.4).

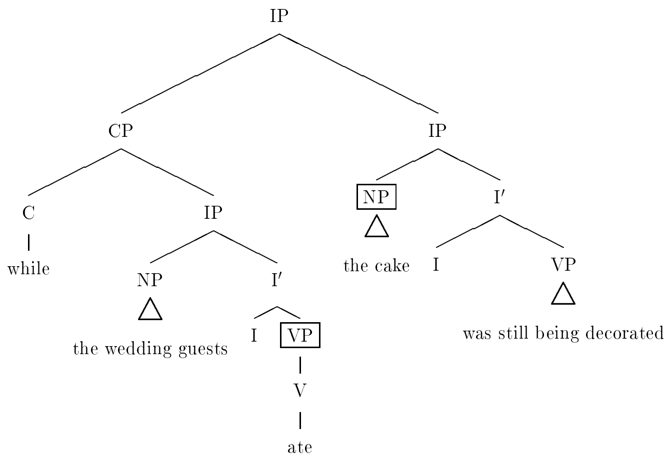
Gorrell explains the difficulty of processing (2.4) in terms of the fact that monotonicity is not preserved in primary relations. In (2.3), by contrast, although secondary relations have to be altered during the parse, primary relations can be built up monotonically, and the sentence is therefore predicted to be easily processable. Below we give a brief description of this, though for more detail and a range of further examples, the reader is referred to Gorrell's original work (1995b).

Consider (2.4). At the point where the parser has just received *the cake*, this NP will have been attached as the direct object of *ate*. Thus, the set of relations will encode the fact that a VP dominates this NP (in the following diagram, we box these two nodes for clarity):



{...,dom(VP,NP),...}

However, the following word, *was* forces a reinterpretation in which the NP *the cake* appears in the matrix clause.

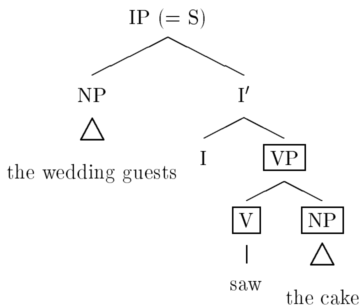


In this revised structure, it can be derived that the original VP now precedes the NP (through the inheritance condition), but this leads to a contradiction, because we now have

$$\text{dom(VP,NP)} \wedge \text{prec(VP,NP)}$$

against the exclusivity condition, and thus the parse is predicted to be impossible.

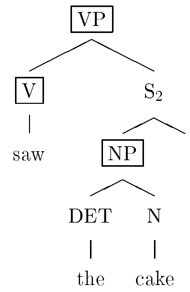
Now consider (2.3). At the point where *the cake* has just been parsed, the tree-description will include an NP dominating *the cake*. This NP is dominated by VP and preceded by V:



The primary relations will include the following:

{..., dom(VP,NP), prec(V,NP),...}

Among the secondary relations, we will have, for example, the assignment of a thematic role to the direct object NP by the verb. Now, encountering the verb *was* will force a reinterpretation to a complement clause analysis, and a consequent revision of this assignment of thematic roles. However, the set of primary relations can be updated monotonically, with the addition of a new S-node (call it S_2).[11] We show the relevant section of the corresponding phrase marker below:



This is achieved by adding the following relations:

{dom(VP,S_2), prec(V,S_2), dom(S_2,NP)}

The addition of the above new relations does not falsify any of the relations carried over from the previous state. For example, both dom(VP,NP) and prec(V,NP) are still true.

In summary, the D-theory formal machinery has allowed Gorrell to isolate a configurational level of representation ("primary relations") which can be built up monotonically, while allowing non-monotonicity at other levels.

3.3.3 Non-monotonic Semantic Interpretation

Gorrell's original model, in common with approaches such as Pritchett (1992), assigns great importance to the syntactic level of representation in explaining garden path phenomena, and as such, does not concern itself with issues of semantic interpretation. The implementation which we discuss in the next chapter inherits this concentration on syntax. Ultimately,

[11] In standard GB theory, and in Gorrell's model, we would also require a new CP node immediately dominating the embedded S (or IP) node. This will mean that the secondary relations of case-assignment and government hold between V and NP before but not after reanalysis. The CP node has been omitted for clarity of exposition.

however, for a processing model to be plausible, we must show that the parser is capable of building semantically interpretable structures. We believe that the best approach for Gorrell's model, as well as the model described in the next chapter, is that suggested in the original D-theory paper (Marcus, Hindle, and Fleck (1983)), in which the "default tree" entailed by the description is created at each state (that is after each word has been attached), by making the "closed world" assumption. We assume that it is possible to build a semantic interpretation from such structures.[12] The non-monotonicity of secondary relations suggests that semantic interpretation must also be non-monotonic, as it is usually assumed that secondary relations, such as theta assignment, are necessary for interpretation. There is no guarantee, for example, that a semantic representation available at a particular parse state will entail any representation at a subsequent state. Consider the following sentence for example:

(3.5) John believed the politician was lying.

The interpretation available in a "snap-shot" taken immediately after *the politician* has been read, on NP object reading, certainly does not entail the interpretation of the whole sentence; in fact, one interpretation is contradicted by the other.

This type of model, in which syntax is given a fairly privileged place in the processing architecture, may be contrasted with models which are concerned primarily with incrementally extracting semantic representation, and in which logical forms are built directly, without the explicit construction of a purely syntactic level of representation (e.g. Pulman (1986), Milward (1994), (1995)). Such models derive logical forms entirely non-destructively by using logical devices such as higher-order abstraction.

3.3.4 Connectedness and Incrementality

The mere existence of garden path phenomena shows that the human parser is incremental in the sense that it does not wait for disambiguating information before committing itself to an analysis of a locally ambiguous material. However, given the incremental nature of syntactic

[12] This view of semantic interpretation is similar in spirit to certain computational approaches, such as that of Shieber and Johnson (1993), which uses Tree Adjoining Grammars (Joshi, Levy, and Takahashi, 1975) to pack together trees which share the same recursive structure (we will discuss TAGs in some detail in the next chapter, in section 4.3.3). In this system, a semantic tree is built up simultaneously with the syntactic tree in the Synchronous TAG formalism. Default semantic values can be obtained from the semantic tree at each point in processing by assuming that no further TAG adjunctions are to be performed.

processing, we are faced with a further question: How fine-grained are the units of incremental processing? In response to this question, many researchers assume a head-driven architecture, in which commitment to a syntactic analysis may only be made when a licensing head has been found in the input (Abney (1987; 1989), Pritchett (1992)). Such a strategy implies that in a head-final language, such as Japanese for example, the processor waits until the final word of a phrase before building that phrase and thus committing itself to an analysis. However, there is experimental evidence from Dutch (Frazier (1987)) and intuitive evidence from Japanese (Inoue and Fodor (1995)) that in such languages, structuring *can* and *does* occur before the head has been encountered:

Consider the following example (Inoue (1991), p.102):

(3.6) Bob ga Mary ni [$t_{nom/i}$ ringo wo tabeta] inu$_i$ wo ageta.
Bob NOM Mary DAT apple ACC eat-PAST dog ACC give-PAST
"Bob gave Mary the dog which ate the apple "

Comprehenders report a "surprise" effect on reaching the first verb, *tabeta* ("ate"). This is explained on the assumption that the nominative, dative and accusative arguments ("Bob", "Mary" and "the apple"), are initially postulated as coarguments of the same clause, in advance of reaching the verb. On reaching the transitive verb "ate", this analysis is falsified, since this verb cannot take a dative argument. However, if, as the head-driven models would predict, the arguments are not structured in advance of the verb, but are held in some form of local memory store, then we have no simple explanation for the surprise effect. Observations such as these have recently been confirmed in a number of experimental studies on head final constructions (Bader and Lasser (1994), Hemforth, Konieczny, and Scheepers (1994a), Yamashita (1994)), all of which provide evidence for attachment commitments before the head of a phrase is read.

Gorrell uses examples such as these to motivate the principle of *Incremental Licensing* (1995b):

> **Incremental Licensing:** The parser attempts incrementally to satisfy the principles of grammar.

However, no formal specification is made of what actions the parser takes if it is unable to satisfy this principle. The crucial question is that of whether the parser should be allowed to buffer constituents. For example, if the current word cannot be attached into the description under construction so as to guarantee a grammatical continuation, is it permissible to keep the word and its associated superstructure in a buffer or stack until the issue is resolved? This question is related to the issue of connectedness in that the size of the stack accessible to the parser corresponds to the amount of structure which may be left unconnected in the parser's memory (see Stabler (1994a), (1994b)) for a discussion of this issue). Gorrell does not provide a discussion of this, though implicitly he does allow the parser to retain unconnected material in its memory, in cases where there is insufficient grammatical information to postulate a structural relation between two constituents. One example of this is the occurrence of a sequence two non case-marked NPs, as in the following centre-embedded example:

(3.7) The man the report criticised was demoted.

On this question, Gorrell claims that "there is no justification for asserting any relation between the two NPs" Gorrell (1995b). This implies that the parser is able to store unconnected material in its memory. If this is so, then it is necessary to constrain the conditions under which material may be added to this store. In particular, if the parser is permitted to shift material onto a stack whenever it fails to make an attachment, then, in a garden path utterance, the error will not be recognised until the end of the input has been reached, when the parser will be faced with a stack full of irreducible structures. This was essentially the problem faced by Abney, whose (1987) model required the use of an unbounded stack. This was because attachment could only be made under a head-driven form of *licensing*, so that in left-branching structures, where the licenser was still unread in the input, it was essential to shift structure onto the stack until the licenser was found, in order for the attachment to be made. The problem was that the parser could not tell whether to continue adding material to the stack, in the expectation of a licenser later in the input, or whether to abandon the parse. In a later version (Abney (1989)), this problem was solved with the addition of LR-states, to indicate whether or not a grammatical continuation could be expected at the current input.

In the implemented model which we introduce in the next chapter, we do not allow the parser to store unconnected material in its memory, insisting instead on *full connectedness*. A more general solution to the problem of fully connected parsing will be discussed in chapter 5.

3.4 Summary

In this chapter, we have given an overview of the *monotonic* processing models, considering in particular, the contribution made to this class of models by Gorrell (1995b). By introducing incremental processing, limiting underspecification to a minimum, and applying monotonicity to only a *subset* of linguistic relations, Gorrell creates a system which is considerably less "minimally committed" than previous models of the deterministic tradition. In Gorrell's model, monotonicity is more properly described as a constraint on possible reanalysis than a way of limiting the initial commitments that the parser makes. The models which we introduce in the remainder of this thesis will inherit this property. However, while Gorrell's model makes an important contribution to the class of monotonic models, it makes very little consideration of the actual processes which are involved in reanalysis. In terms of Marr's (1982) levels of explanation for theories of mental processes, Gorrell's model may be seen as a *computational level* theory, in that it specifies the logic of the relevant computation. In the next chapter, however, we will go one step further, and introduce a model which may be seen as an *algorithmic* level implementation of Gorrell's theory, in that it specifies the details of *how* the relevant computation is carried out. We will argue that it is important for a theory of reanalysis to consider these issues, and show how such algorithmic considerations can lead to novel predictions.

Chapter 4

Monotonicity and Constituent Structure

4.1 Introduction

4.1.1 D-theory and Psychological Models

As we have seen, one of the aims of theories of reanalysis is to capture differences in processing difficulty associated with different types of reanalysis.[1] From a theoretical point of view, the attraction of the monotonic D-theory based models such as Gorrell (1995b), which we discussed in the previous chapter, is that they offer the possibility of deriving the required constraints on structural revisions "for free", as a consequence of the representational formalism used by the parser, removing the need to stipulate the constraints separately. By the same token, however, these models suffer from a conceptual problem, which becomes particularly clear when one attempts to provide a computational implementation. The problem can be summarised as follows. As we argued in chapter 2, the existence of constraints on the class of possible structural revisions may be assumed to be a consequence of systematic strategies adopted by the human parser to cope with the inherent computational complexity of the search involved in the reanalysis task. However, the mere adoption of the D-theory representational formalism in no way guarantees that the search mechanisms used by the parser will be limited in the required manner. For example, in D-theory based models, it

[1]The material reported in this chapter is largely drawn from Sturt and Crocker (1996).

is predicted that reanalysis involving the *addition* of structural relations (e.g. precedence and dominance relations between nodes) is more difficult to achieve than reanalysis which requires the *deletion* of structural relations. However, this still leaves the question of defining the search strategies by which the parser decides *which* structural relations to add at each parse state.

Consider, for example, sentence 2.3, repeated below:

(2.3) The wedding guests saw the cake was still being decorated.

Assuming again that *the cake* is initially attached as the direct object of *saw*, then at the point where *was* is received, a number of structural relations must be added, to incorporate a new clausal node into the structure with *the cake* as its subject (see page 52 for a discussion of how adding these relations results in a reanalysed structure). The question, then, is how the processor chooses which relations to add. Clearly, an algorithm which simply proposed sets of new structural relations at random, and subsequently filtered out those which were inconsistent with the theory of trees, would be computationally implausible, representing a totally unconstrained and possibly non-terminating search procedure, seemingly at odds with the idea that the constraints on structural revisions are computationally helpful to the human language processing mechanism. Instead, we believe it is much more plausible to assume that the processor uses systematic strategies in deciding which relations to add. There are a number of possible ways in which such strategies can be defined. Given the existence of systematic strategies, or search procedures, as we will refer to them in this chapter, we can still imagine a number of possible architectures in which such procedures could be instantiated. For example, the processor could use some systematic heuristics to propose possible new sets of relations, which are, again, filtered out with reference to the theory of trees. However, a second possible architecture is one in which the choices of the parser are limited to a constrained set of parsing operations, which are guaranteed to preserve D-theory coherence, and the parser is simply incapable of revising in such a way as to violate the D-theory conditions. Such a processing regime is likely to be more computationally efficient than one in which solutions are initially proposed, and subsequently filtered through a D-theory "clash check"[2] The approach is also interesting on a methodological level, since the explicit statement of restricted parsing

[2]Though, we note that the "clash check" itself may be performed in polynomial time (Cornell, 1994).

operations and search strategies for their application is likely to result in a more restrained, and thus more predictive, model.

In this chapter, we will take this stronger approach. Specifically, we will describe a model which builds representations that preserve the D-theory coherence conditions, but where this constraint is obtained as the result of restricting the operations available to the parser, rather than imposing a filter on proposed output. In particular, we propose that the parser proceeds by incrementally assembling descriptions of lexically anchored tree fragments, and that it has at its disposal two composition operations, *simple attachment*, which identifies the root of one tree with a node on the fringe of another, and *tree lowering*, which inserts one tree inside another at an intermediate point on the right frontier. These operations are related to the operations of *substitution* and *adjunction* found in tree adjoining grammars (henceforth TAGs) (Joshi, Levy, and Takahashi (1975), Schabes, Abeillé, and Joshi (1988)), though we will see that an important difference between TAG adjunction and the tree-lowering operation means that the latter, but not the former, can be used for reanalysis.

As we mentioned in the previous chapter, there has been a discernible trend towards constraining the power of the original D-theory model in order to sharpen its psychological plausibility. In the implementation described in this chapter, we take the constraints of Gorrell's model as a starting point, but, by limiting the operations of the parser, we derive a parser which is more restrictive than Gorrell's, in the sense that only a subset of attachments licensed in Gorrell's model is licensed in the model proposed here. We believe that this course of action has the following advantages over previous D-theory based models:

1. By limiting the attachment operations available to the parser, we directly address the search problem; difficult reanalysis is unavailable simply because it is not in the parser's repertoire, and not because it is filtered out as part of a generate-and-test procedure. This has advantages for the computational efficiency of the parser, but also yields a model which is more restrictive, in the sense that only a subset of the D-theory valid attachments/reanalyses are possible. This gives the model greater predictive power.

2. The explicit definition of a set of parsing operations, which are related to a well-understood grammar-formalism (Tree Adjoining Grammar), makes the model relatively simple to implement computationally.

3. Finally, psycholinguists have recently begun to give serious consideration to the question of *how* the parser reanalyses (Fodor and Inoue, 1994), as opposed to *what structures* the parser can and can't reanalyse. This implies a need for computational models capable of making explicit predictions on this question. The formulation of parsing operations, and search strategies for their application, as we provide in this chapter, represents a contribution to such a research programme.

Before we proceed with the discussion, we should point out one important consequence of the restrictive architecture which we describe in this chapter. Since the parser is only capable of monotonic reanalysis, it must be the case that *non*-monotonic reanalysis is achieved through a separate mechanism. This is essentially the view of Pritchett (1992) and Abney (1989), who both propose that "hard" reanalysis is achieved outside the "automatic" human sentence processor, using higher level reasoning. A less restrictive view of the monotonicity hypothesis would be to see monotonicity as a *preference* rather than as an absolute *requirement*, so that, for example, the parser would attempt to reanalyse monotonically before considering non-monotonic reanalysis solutions, but would be *capable* of non-monotonic reanalysis. In either case, a full model would have to consider how non-monotonic reanalysis is achieved, since systematic effects in non-monotonic reanalysis are known to exist, and these effects require explanation.[3]

4.2 Constraints on the Model

The implementation described here obeys the following constraints, which are intended to capture, and build on, the conditions described in Gorrell's work. In particular, informational monotonicity is defined, as well as full specification. The condition of incrementality is stronger than that implied by Gorrell, since it insists on full connectedness. that is to say that at any stage, the parser has access only to a single set of relations describing a fully connected tree, and each word has to be incorporated within this structure as it is encountered.

Here, then, are the constraints which apply to our implemented model:

1. **Strict Incrementality:** Each word must be connected to the current tree description

[3]For example, Ferreira and Henderson's (1991) head position effect (described above in chapter 2) involves what would be considered as non-monotonic reanalysis in Gorrell's (1995b) model.

at the point at which it is encountered through the addition of a non-empty set of relations to the description.

2. **Structural Coherence:** At each state, the tree description should obey the conditions on trees:
   (a) Single Root Condition.
   (b) Exclusivity Condition
   (c) Inheritance
3. **Full Specification of Nodes:** Tree-descriptions are built through the assertion of dominance and precedence relations between fully specified nodes. In the current implementation, each node is a triple ⟨Cat,Bar,Id⟩, consisting of category *Cat*, bar-level *Bar* and an identification number *Id*. Each of these three arguments must be fully specified once the structure has been asserted.[4]
4. **Monotonicity:** The tree-description at any state n must *be a subset* of the tree-description at state $n+1$.[5] Thus the parser may not delete relations from the tree description.
5. **Obligatory Assertion of Precedence:** If two or more nodes are introduced as sisters, then precedence relations between them must be specified.
6. **Grammatical Coherence:** At each state, each local branch of the phrase marker described must be well-formed with respect to the grammar.[6]

[4]As an anonymous reviewer has pointed out, the insistence on full specification of syntactic category is almost certainly too strong; it does not allow us to capture the cases in which the part of speech of a word is ambiguous, and resolved only after subsequent input has been read, with little appreciable processing cost (Frazier and Rayner (1987)). Though we have not worked out the details, a fuller account might allow category names to be changed, while retaining certain aspects of the basic structural skeleton, or to introduce parallelism or further underspecification for lexical category ambiguity. These solutions would, however, not follow directly from the D-theory formalism.

[5]In fact, since (1) requires the set of relations added to the description at each word to be non-empty, the description at n is a *proper* subset of the description at $n+1$.

[6]Note that adopting the Grammatical Coherence constraint does not render Strict Incrementality vacuous. Strict Incrementality requires full connectedness, and no buffering of input, while Grammatical Coherence is intended to guarantee that the (connected) 'default tree' produced at each incremental point in processing is specified enough to support semantic interpretation, at least in principle.

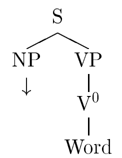
4.3 Coherence-Preserving Parsing Operations

4.3.1 Syntactic Representation

The syntactic representation assumed is similar to that used in Tree-Adjoining Grammars (henceforth TAGs) (Joshi, Levy, and Takahashi (1975)), and in particular, *Lexicalised* Tree Adjoining Grammars (Schabes, Abeillé, and Joshi (1988)). Each lexical category is associated with a set of structural relations, which determine its *lexical subtree*. For example, the verb category for English contains the following relations:

{dom(VP,V^0), dom(V^0,Word), dom(S,VP),
dom(S,NP),
prec(NP,VP)}

These define the following subtree, where we borrow a tradition from the TAG literature and represent an attachment site with a downward-pointing arrow.[7]



As each word is encountered in the input, the parser attempts to add the set of structural relations defining that word's lexical subtree to the set of structural relations defining the global representation under construction, in a way which is similar to the *Assertion Set* approach to parsing (Barton and Berwick, 1985). There are two possibilities for this: simple attachment and tree-lowering, corresponding roughly (though not exactly, as we shall see later) to substitution and adjunction in TAG.[8]

[7]Note that, strictly speaking, this is the minimal, or "default" tree defined by the set of relations given above. We will often speak of descriptions as though they are fully specified trees when we believe this to be clearer.

[8]Note that the parser is not *head-driven*, since, for example, we allow lexical subtrees to include material not dominated by the word's maximal projection. In a head-final construction, this allows arguments and adjuncts to incorporate themselves into the projection of the licensing head before this head is reached in the input. Empty heads do not play any structure building role in the parser.

4.3.2 Simple Attachment

The parser is capable of performing simple right and left attachment, illustrated schematically in Fig.4.1. A lexical subtree may contain attachment sites to the left or the right of the word from which it is projected. These are distinguished empty nodes, which must be filled in accordance with the conditions on trees (the parser effectively keeps these nodes on a stack; see appendix for details). Intuitively, left-attachment consists in attaching the current global tree onto the left corner of the subtree projection of the new word, while right attachment consists in attaching the subtree projection of the new word onto the right corner of the current global tree. This is done by simply identifying the root node of one subtree with an attachment site from the other subtree, and adding the required structural relations to the global representation. The attachment operations are similar to Abney's (1987; 1989) *Attach-L* and *Attach* respectively. In the following definitions, we use the term *current tree description* to refer to the set of relations describing the global phrase-marker currently in the parser's memory, in other words, the parser's *left context*. The term *subtree projection* is used to refer to the set of relations corresponding to the lexical category of the new word encountered in the input.

Left Attachment:
Let D be the current tree description, with root node R. Let S be the subtree projection of the new word, whose left-most attachment site, A is of identical syntactic category with R. The updated tree description is $S \cup D$, where A is identified with R.

Right Attachment:
Let D be the current tree description, with the first right attachment site A. Let S be the subtree projection of the new word, whose root R is of identical syntactic category with A. The updated tree description is $S \cup D$, where A is identified with R.

The parser is also capable of creating a new attachment site with reference to a verb's[9] argument structure. For example, if a transitive verb is found in the input, then a new right attachment site is created for a NP, and a new NP node is "downwardly projected" as a sister to the verb.

So, for a simple transitive sentence such as *Polly eats grapes*, first *Polly* is projected to a

[9] This is also done with other argument assigners, such as prepositions.

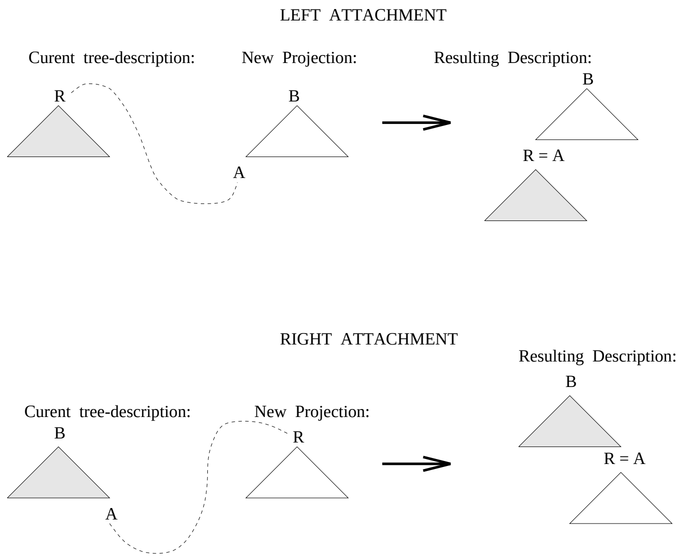


Figure 4.1: Left and Right Attachment

NP and instantiated as the current global tree. This NP will match the left attachment site of *eats*, and so left attachment will be performed. Since *eats* is a transitive verb, a new NP attachment site will be created, and this will be suitable for right attachment of the projection of *grapes*.

It should be noted that the use of downward projection requires a systematic method for choosing amongst alternative subcategorization frames in cases where the verb is ambiguous. Gorrell (1995b) defines the notion of simplicity: "No vacuous structure building" — the principle which is used to explain the initial preference for NP attachment in the case where a verb may subcategorise either for a NP or for a clause. In the case of an optional argument, however, the processor cannot employ downward projection, since there is no guarantee whether or not the argument will appear in the input. This implies that, if the parser is able to cope with the attachment of optional arguments, it must be capable of attaching an argument without first downwardly projecting that argument. One option for dealing with this, which we will use in the implementation described in chapter 6, is to project complements "on demand", when their presence becomes indicated by the input.

If simple attachment is not possible, then the parser attempts to perform a second mode of attachment, *tree-lowering*. Before describing tree lowering, we will briefly review the adjunction operation of TAGs, since the two operations share some common features.

4.3.3 TAG Adjunction

A TAG contains a set of elementary trees, which are divided into *initial* and *auxiliary* trees. An initial tree is rooted in a distinguished *start symbol* non-terminal node, and has terminal symbols along its frontier. An auxiliary tree is rooted in a non-terminal node, and its frontier includes a further non-terminal node, which must be of the same type as the root. This non-terminal node is known as the foot node.

Most versions of TAG allow two operations for combining one tree with another; substitution and adjunction. Substitution is essentially the same operation as simple attachment as defined above. To explain adjunction, we refer to an example instantiation given in Fig.4.2.

The following definition is adapted from Joshi, Vijay-Shanker, and Weir (1991). Let α be be an initial tree containing a node n with category c (this is the node marked 'c:n' in Fig.4.2).

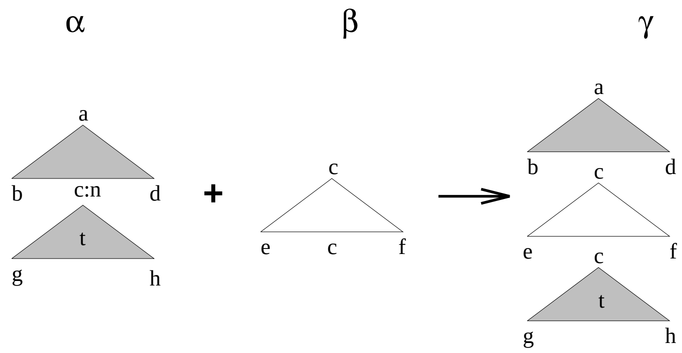


Figure 4.2: TAG adjunction.

Let β be an auxiliary tree, whose root and foot nodes are also of category c. The adjunction of β to α at node n will yield the tree γ that is the result of the following operations:

1. The subtree of α dominated by n is excised. We call this excised subtree t.
2. The auxiliary tree β is attached at n, and its root node is identified with n.
3. The excised subtree t is attached to the foot node of β, and the root node, n of t is identified with the foot node of β.

It is simple to define a version of TAG adjunction in terms of sets of structural relations which describe trees. The trees α, β and γ can be described by sets of relations, which we call A, B and ? respectively. For example, A (which describes α), is as follows — (Again, we use `c:n` to represent the node n with category C)

$A = \{$`dom(a,b), dom(a,c:n), dom(a,d), dom(c:n,g), dom(c:n,h), prec(b,c:n), prec(c:n,d), prec(g,h)`$\}$

The set B is similarly defined. Now, to adjoin β to α, we find the set of local relations $L \subseteq A$ in which `c:n` participates:[10]

[10] The 'local relations' in which a node N participates at state S are those dominance and precedence relations which define the mother and sisters of N at S .

$L = \{$`dom(a,c:n), prec(b,c:n), prec(c:n,d)`$\}$

We then build a new set of relations N, which is L with all occurrences of `c:n` replaced by the root node of β (call it `c:r`)

$N = \{$`dom(a,c:r), prec(b,c:r), prec(c:r,d)`$\}$

We then identify the foot node of β with `c:n`, so that B consists of the following set of relations:

$Y = \{$`dom(c:r,e), dom(c:r,c:n), dom(c:r,f), prec(c:n,f)`$\}$

Now ? is simply defined as follows:

$? = A \cup N \cup B$

As the reader can verify, the derived set, ?, describes the tree γ, as required.

4.3.4 Tree-Lowering

In standard TAG adjunction, the auxiliary node (that is, the node at which adjunction is performed, c:n in the above example) is, informally speaking, 'split' into a bottom half, which is identified with the foot node of the auxiliary tree, and a top half, which is identified with the root node of the auxiliary tree. However, in the version of adjunction which we have defined above, in terms of sets of relations, we identify c:n only with the foot node of the auxiliary tree, and treat the root node of the auxiliary tree (c:r in the above example) as a separate node that takes over the place previously occupied by c:n. One consequence of this is that we are able to generalise the adjunction operation to include cases where the root and foot nodes are of distinct syntactic categories, which is impossible in a standard TAG.[11] It is this property which allows us to capture reanalysis phenomena in a monotonic fashion, since it provides a way of replacing node X with node Y, while also guaranteeing that the original position of X dominates the revised position of X. Consider example 2.3, repeated below.

(2.3) The wedding guests saw the cake was still being decorated.

Schematically, what we need in the above case is illustrated in Fig. 4.3.

[11]Vijay-Shanker (1992) discusses a version of TAG in which root and foot nodes need not be of the same category. However, Vijay-Shanker's system differs from that proposed here in that it explicitly divides auxiliary nodes into a top and bottom half, into which structure can subsequently be inserted, and is intended to capture grammatical, rather than processing phenomena. (See Sturt and Crocker (1995) for a fuller discussion).

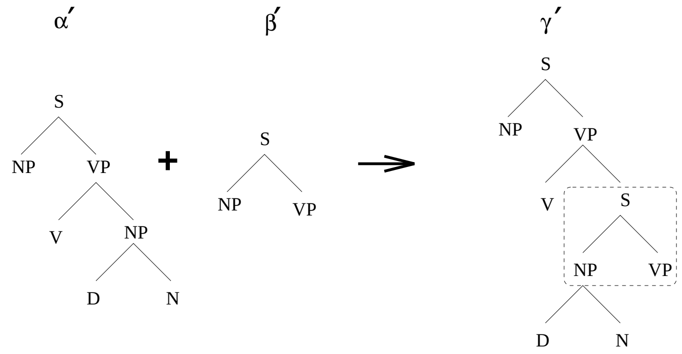


Figure 4.3: Reanalysis as the insertion of one tree inside another. The inserted material is enclosed inside the dotted line.

Here, the nodes corresponding to root and foot nodes of the auxiliary tree are the S node and NP (subject) node of the embedded clause respectively. In order to accommodate β' into α' in the desired monotonic fashion, therefore, we will have to drop the requirement for root and foot nodes to be of identical syntactic category. However, we must constrain this operation so that it may only be employed in cases where the root node of the auxiliary tree is licensed in its new adjoined position. In the above case it *is* licensed, since *know* may subcategorize for a clause.[12]

In order to maintain structural coherence, the new word attached via tree-lowering must be preceded by all other words previously attached into the description. We can guarantee this by requiring the lowered node to dominate the last word to be attached. We also need to ensure that, to avoid crossing branches, the lowered node does not dominate any unsaturated attachment sites (or "dangling nodes"). In other words, in order to obey the theory of trees, the node selected for tree-lowering must be accessible in the following sense:[13]

DEFINITION **Accessibility:**

Let N be a node in the current tree description. Let W be the last word to be

[12]In the current implementation, lexical information for the verb *know* is re-accessed in order to find the alternative subcategorization frame.

[13]Though Accessibility can be derived from the theory of trees, the actual parser does not reason from first principles in order to find accessible nodes, but explicitly uses the above definition. This is done for efficiency reasons.

attached into the tree.

N is accessible iff N dominates W, and N does not dominate any unsaturated attachment sites.[14]

Note that it is not necessarily the case that all nodes on the right frontier of the tree are accessible, nor that all accessible nodes are on the right frontier of the tree. For example, a "dangling node" will not dominate any lexical material, even though it might be on the right frontier of the tree, and therefore it will not be accessible. Also, there may be a node which dominates the last word to be attached, and which is therefore accessible, but which precedes a dangling node, and is therefore not on the right frontier of the tree.

The tree-lowering operation is defined as follows, and illustrated diagrammatically in Fig. 4.4:

DEFINITION **Tree-lowering:**

Let D be the current tree description. Let S be the subtree projection of the new word. The left attachment site A of S must match a node N accessible in D. The root node R of S must be licensed by the grammar in the position occupied by N. Let L be the set of local relations in which N participates. Let M be the result of substituting all instances of N in L with R. The attachment node A is identified with N.

The updated tree-description is $D \cup S \cup M$.

Note that, in order to check whether the root node R of the new subtree projection is licensed in its new position, it may be necessary to access subcategorization information associated with a word long past in the input, and which is no longer *accessible* in the sense defined above. This is the case in the above example, where the subcategorization frame of *knows* has to be checked to allow the attachment of the clausal node as its sister, although the V^0 node of *knows* itself is no longer accessible. It would be interesting to investigate how far inside the tree the parser is capable of looking in order to extract this type of information.[15]

The parser is constructed in such a way that, if at any point in the parse, simple attachment fails, the accessible nodes of the current tree-description are considered until a node is found

[14]Note that Tree-lowering constrained by Accessibility results in a system which is very similar to that of Stevenson's competitive attachment model (1993; 1994a). See section 4.8 for discussion.

[15]We are grateful to Martin Pickering for bringing this point to our attention.

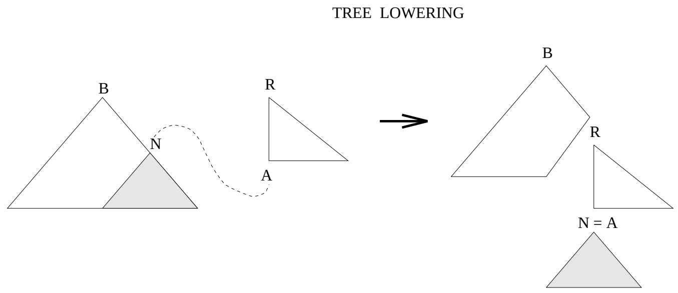


Figure 4.4: Schematic illustration of Tree Lowering. The node R must be licensed in the position previously occupied by N

at which tree-lowering may be applied. Note that, tree-lowering can capture many effects of standard TAG adjunction, and it is therefore possible to use this operation for the attachment of post-modifiers, which is the course of action taken in this implementation. The preference for argument over adjunct attachment (Abney, 1989; Clifton, Speer, and Abney, 1991) is captured by the fact that tree-lowering is only attempted in cases where standard attachment fails.[16] Note, however, that we do not claim that adjunct attachment exhibits the same cost as reanalysis. The distinction is that adjunct attachment involves only *adding* a new dependency, while reanalysis crucially involves *breaking* a dependency[17] (see section 4.7 for a discussion of this issue).

4.4 Top-Down Prediction

We mentioned earlier that, (for head-initial languages like English), when the parser encounters a head requiring a following internal phrasal argument, this argument is projected top-down and asserted as a so-called "dangling node". However, there will be cases where the word immediately following the head cannot be directly connected to such a dangling node. Consider the following example:

(4.1) Mary thinks John .....

[16]In Gorrell's original model, this is accounted for by the Principle of Simplicity (Gorrell (1995b)).

[17]In Gorrell's (1995b) terms, reanalysis involves deleting secondary relations.

On encountering the verb *thinks*, the parser must project a clausal node, since this verb can only subcategorize for a clause. However, the NP *John* cannot be directly connected to this node, since it is of the wrong syntactic type. The current implementation addresses this problem by adopting Crocker's approach (1994; 1996), in which the "functional structure" of the clause (CP, IP) is projected top-down along with the NP subject node in the specifier position of IP. This provides an immediate attachment site for the embedded subject. This is an example of the parser's "non-lexical structure building", as discussed in the appendix.[18]

We will discuss a more general solution to the problem of non-lexical structure building in chapter 6, where connections between the current word and the left context are only built when evidence appears in the input.

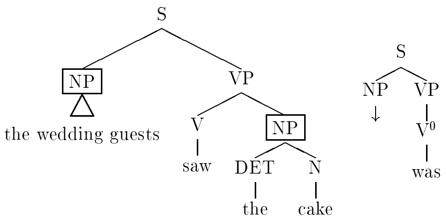
4.5 Search Strategies for Coherence-Preserving Reanalysis

4.5.1 English

Consider again example 2.3:

(2.3) The wedding guests saw the cake was still being decorated.

At the point where standard attachment fails, the parser is faced with the task of incorporating the projection of *was* into its representation. In this case, there is only one accessible node at which tree-lowering may be applied in such a way that grammatical licensing conditions are met, and that is the NP boxed in the diagram below:



In the above, then, the word *was* uniquely disambiguates the local ambiguity. However, there may be occasions where reanalysis does not coincide with unique disambiguation. In particular, the parser may be faced with a choice of alternative lowering sites in much the same

[18]Milward's rule of *state prediction* ((1994), (1995)) gives a more general solution to this problem using a dynamic grammar formalism (see Sturt and Crocker (1995) for further discussion on this point).

way as it is faced with a choice of alternative attachment sites at the onset of a standard local ambiguity. and therefore we must consider the search strategy, or heuristics used to choose between such lowering sites. For example, imagine that the following utterance has just been processed:

(4.2) I know [$_{NP_1}$ the man who believes [$_{NP_2}$ the countess]]

Now, imagine that the utterance continues with the verb *hates*. The verb must be attached via tree-lowering, but now there are two accessible nodes where the operation can be applied; NP_1 and NP_2. It intuitively seems that the lower site, NP_2 is preferred, and we will discuss some experimental evidence supporting this intuition in chapter 7. This can be seen more clearly in the following sentences, where binding constraints force a particular reading; (4.3.a), where lowering is obligatorily applied at the lower NP, is easier than (4.3.b), where lowering can only be applied at the higher NP:

(4.3) a. I know the man who believes the countess hates herself.

b. I know the man who believes the countess hates himself.

In the above examples, though the verb *hates* triggers reanalysis, it is the following reflexive pronoun, *himself/herself* which uniquely disambiguates the structure. If it is indeed the "low" reanalysis (corresponding to (4.3.a)) that is favoured, then in the dispreferred case, (4.3.b), the parser will *mis*-reanalyse on encountering *hates*, only to experience what we might think of as a "second order" garden path effect at the disambiguating signal, *himself*, where the preferred reanalysis is seen to have been mistaken.

If the preference to reanalyse at a low, more recent site does indeed exist, it may be compared with a similar preference to attach post-modifiers to low sites, which follows from from principles such as *Right Association* (Kimball (1973), Phillips (1995), (1996)), or *Late Closure* (Frazier, 1978). Thus, for example, the preferred reading for *John said Bill left today* is (4.4.a), where the adverbial appears in the lower clause, as opposed to (4.4.b), where it appears in the higher clause.

(4.4) a. John said [Bill left today]

b. John said [Bill left] today

Since tree-lowering is also used for post-modifier attachment, we would expect the search strategy used in examples such as (4.3.a) to share some features in common with that in (4.4.a).

A possible strategy which can be used, then, is to search the set of accessible nodes in a bottom-up direction. We define the *current node path* as the ordered set of accessible nodes. To implement the bottom-up search, the current node path is ordered from bottom to top, so that, in the path $\langle N_1, N_2, ..., N_n\rangle$, the node N_1 is the node immediately dominating the last word to be processed (i.e. the lowest accessible node), N_n is the highest accessible node, and N_i immediately dominates N_{i-1} for each pair of adjacent nodes in the path. The parser will then consider, in the order given, each node in path as a possible lowering site.[19]

In the bottom-up search, then, the parser considers the first node in the path, N_1, and attempts to lower. If this is unsuccessful, it moves to the next node, N_2 (i.e. the node immediately dominating N_1), and again attempts to lower. The process continues, with the parser considering successively higher nodes until either lowering is successful, and the parser can move on to consider the next word, or the current node path is exhausted, in which case, the parser fails, and the string is rejected as either a conscious garden path or ungrammatical.

It will be noted that the simple bottom-up search reflects a preference to lower the most recent node possible. However, we recognise that recency is not the only factor in the attachment of post-modifiers (Cuetos and Mitchell (1988), Gibson et al. (1996)), so that the pure bottom-up search we propose here can only be seen as an approximation. In section 4.7 we will discuss possible factors which might influence the selection of a site for tree-lowering.

4.5.2 The Restrictiveness of the Model

The reader can verify that the attachment and lowering operations as defined above are guaranteed to preserve the conditions on trees, and thus to satisfy the monotonicity criterion (Gorrell's *structural determinism*). For the attachment case, this is trivially true. For the lowering case, it can easily be seen that the source position of the lowered node (i.e. its configurational position before the lowering operation is applied) is guaranteed to dominate its target position (*after* lowering has been applied). The result of this is that the present

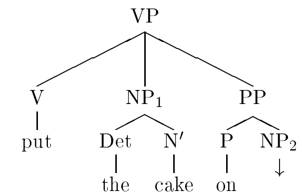
[19]Note that different search strategies may be implemented by changing the ordering function, and this will be exploited in section 4.6 to deal with Japanese.

model is at least as restrictive as Gorrell's model; that is to say that all constructions predicted as conscious garden paths in Gorrell's model are also ruled out in the present model (i.e. they do not receive a parse).

Consider the following well known conscious garden path example:

(4.5) John put the cake on the table in his mouth.

Assume that the first PP *on the table* is initially attached as the location argument of the verb *put*. In the present model, as soon as the preposition *on* is attached under the projected PP node, all nodes preceding the PP are "closed off" from accessibility. In particular, no adjunction within the NP_1 will be possible since the NP_1 will never dominate "the last word to be processed" at this or any other subsequent point in the parse. The result of this is that a second PP (*in his mouth* in (4.5)) can subsequently be attached as a modifier of *the table*, but that the revisions required to recover the plausible reading (i.e. re-attaching the first PP, *on the table* as a modifier of the direct object, *the cake*, and attaching the second PP, *in his mouth*, as the new locative argument of *put*), will not be possible.[20]



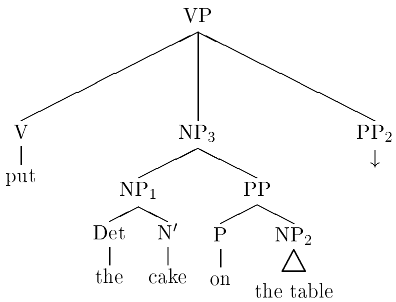
In Gorrell's model, this restriction is captured by the fact that the precedence relation between NP_1 and PP would have to be altered if PP were to be subsequently reanalysed as an adjunct within NP_1 (otherwise NP_1 would simultaneously dominate and precede PP).

As an aside, it is worth noting that in the present model, once PP is attached in its argument position, no adjunction is possible, either to NP_1 itself, or to any node dominated by NP_1, whereas in Gorrell's model, adjunction is ruled out to non-maximal projections *properly dominated*[21] by NP_1, but not to the NP_1 itself (i.e. the maximal projection). The following diagram illustrates the adjunction of PP to the maximal projection NP_1. It can be seen that

[20]This prediction is supported by on-line experimental evidence. Adams (1995) reports that reanalysis of the type required for (4.5), which she calls "theme repair", is more difficult than other options for attaching the second PP, including attachment of the second PP as a modifier of *the table*.

[21]X properly dominates Y iff X dominates Y and $X \neq Y$.

in Gorrell's model, the adjunction can be performed simply by adding a new NP node (which we index NP_3) which dominates NP_1 and PP. This is because the precedence relation between NP_1 and the PP node in its original argument attachment site still holds after the adjunction of PP to NP_1. In contrast, the parsing model outlined in this chapter cannot perform this adjunction once the PP has been attached in its argument position, because, as explained above, NP_1 becomes inaccessible to tree lowering as soon as the argument attachment of PP has been made.



Thus, in order to account for the difficulty of such examples, Gorrell would need to rule out the use of this type of adjunction explicitly, either by appealing to the competence theory, or by placing a constraint on the parser forbidding the "extension" of a maximal projection in this way. This is important, because other well-known garden path examples are predicted to be difficult in Gorrell's model for essentially the same reasons (i.e. because the precedence relation between an NP and a following phrase has to be "converted" to a dominance relation), and would therefore also rely on ruling out this type of adjunction. Examples are *The horse raced past the barn fell* and *The psychologist told the woman he was having trouble with to leave*. We do not necessarily wish to argue for or against the NP adjunction hypothesis on the competence level[22], but rather we use this as an example to show how the model proposed here is essentially more restrictive than Gorrell's in its range of predictions.

Finally, we would like to point out a class of examples for which both Gorrell's model and the model proposed here under-predict. Gorrell (1995b) , ch.5, points out that a model constrained by structural determinism must allow *weak interaction* of non-syntactic knowledge,

[22]Though, the existence of sentences such as *The sweet and the cake on the table looked delicious*, where both NPs *the sweet* and *the cake* seem to be modified by the PP *on the table* make this a reasonable assumption.

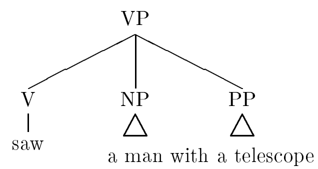
in the sense of Crain and Steedman (1985)). That is to say that, at each point in processing, it must be possible for discourse information and real world knowledge to be used in order to choose among different alternative analyses proposed by the syntactic processor. This means the model faces problems in cases where the implausibility of an initially preferred analysis forces *retrospective* reanalysis. Consider the examples in (4.6):

(4.6) a. John saw the man with a telescope.

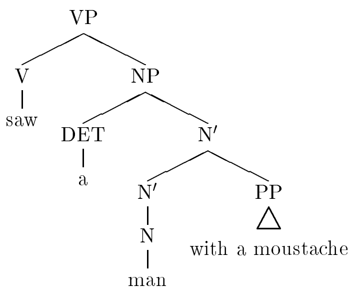
b. John saw the man with a moustache.

Intuitively, there is no serious difficulty associated with either the instrumental reading, in which the PP is attached as an argument of the verb, or with the analysis in which the PP is a modifier of the noun. However, whichever of these analyses is initially chosen, monotonic reanalysis to the alternative analysis is not possible. For example, assume that there is an initial preference for the instrumental reading, in which the PP is attached as a daughter of the VP (Rayner, Carlson, and Frazier, 1983). The constraint of incrementality means that this attachment of the PP must be made as soon as *with* is encountered. In the case of (4.6.a), this analysis remains plausible, since a telescope is a reasonable instrument with which to see a man. In the case of (4.6.b), however, the instrumental analysis is eventually found to be implausible, since moustaches are, in general, rather poor optical instruments, and *with a moustache* will have to be reattached as an adjunct of the NP *a man*. We show the two analyses in (4.7) and (4.8).

(4.7)



(4.8)



On encountering the word *moustache* in (4.6.b), the parser will have to reanalyse from the structure in (4.7) to that in (4.8). But this will result in an incoherent description, since, in (4.7) we have prec(NP,PP), and in (4.8) we have dom(NP,PP), against the exclusivity condition, and yet, though on-line preferences are known, neither of the two sentences given in (4.7) or (4.8) seems to cause conscious processing difficulty. Similarly, if the noun modifier attachment is initially made, reanalysis to the instrumental reading will be impossible due to the necessity to break a dominance relation between the NP and the PP.[23] The problem is mirrored in the present model by the fact that, in the instrumental reading, the NP along with all the nodes it dominates, become inaccessible for tree-lowering as soon as the preposition *with* is attached.

In chapter 5, we will discuss this problem further, and offer a solution in terms of a model which is monotonic at the level of *thematic* structure.

4.6 Processing Japanese

In the previous section, we saw that, at the point of reanalysis, the parser may be faced with a choice of possibilities at which the lowering operation may be applied, necessitating the definition of a search strategy. In this section we will look at a class of examples involving the reanalysis of relative clauses in Japanese, where just such a choice is found.

We will firstly look at Gorrell's explanation for the distinction between conscious and unconscious garden paths for this type of example in terms of a comparison between the set of relations describing the phrase marker at two snap-shots of processing. We then show

[23]The reader can verify that a binary left- branching variant of (4.7) will suffer the same problem.

how this "static" explanation fails to make an adequate distinction between the easy and hard cases of reanalysis. We subsequently demonstrate the important role of the search strategy in explaining this distinction, showing, incidentally, how the bottom-up strategy we proposed to account for the English data predicts the opposite of the observed results. We then give a (speculative) proposal of a top-down, weakly interactive search, which allows a more satisfactory explanation.

4.6.1 Grammatical Assumptions

Before we move on to a discussion of the processing data, we will firstly consider the assumptions which underly the implementation of the Japanese grammar.

We are assuming an analysis of Japanese grammar which employs many features of the model of Korean and Japanese morphology and syntax made by Sells (1995) and Cho and Sells (1995), though, for practical purposes, the grammar is considerably simplified in comparison with these proposals, and lacks their full coverage.

We assume, firstly, that an argument of a verb, such as a case-marked NP, or a complement clause (which will be marked with an obligatory complementizer), may "select" the verbal projection which will be its licensing head.[24]

Furthermore, we assume that lexical categories in Japanese project to no higher bar level than X′ (Sells (1994)).[25]

As far as head-final parsing is concerned, the effect of this is that case marked noun phrases, as well as other potential arguments of a verb, may attach themselves into a verbal projection (effectively by adjoining at the V′ level), before the head verb of that projection has been read. This allows the parser to maintain a fully connected representation when assembling the arguments of a head final construction. Although this method depends on certain features of grammar which may not hold of all head final languages, we will consider more general solutions to the problem of connectedness in chapter 6.

[24]This reversal of the direction of selection between head and argument is reminiscent of the *Type Raising* rule, a theorem of the Lambek Calculus (Lambek, 1958). The type raising rule has been used in categorial grammar for modelling a number of linguistic phenomena (see Dowty (1988), for example).

[25]In our discussion of Japanese, label "S" is to be interpreted as "the highest V′ node of a verbal projection".

4.6.2 The Processing Data

The issues we will be discussing concern constructions such as the following:[26]

(4.9) a. [Mary ga sinseihin wo $t_{loc/i}$ kaihatusita] kaisya$_i$ ga tubureta.
Mary NOM new product ACC developed company NOM went bankrupt

"The company where Mary developed the new product went bankrupt."
(Inoue (1991))

b. Yamasita ga [$t_{nom/i}$ yuuzin wo houmonsita] siriai$_i$ ni tegami
yamasita NOM friend ACC visited acquaintance DAT letter
wo kaita.
ACC wrote
"Yamasita wrote a letter to an acquaintance who visited his friend."
(adapted from Mazuka and Itoh (1995))

c. Yamasita ga yuuzin wo [Ø$_{nom}$ $t_{acc/i}$ houmonsita] kaisya$_i$ de mikaketa.
Yamasita NOM friend ACC visited company LOC saw

"Yamasita saw his friend at the company he visited."
(adapted from Mazuka and Itoh (1995))

In all of the examples in (4.9), we assume that a clause is initially built containing an (overt) subject, object and transitive verb. However, subsequent appearance of the noun shows that this clause cannot be the main clause. Therefore the parser must find some way of connecting the noun with the clause. The most direct way of doing this is to reinterpret the clause, or part of it, as a relative clause modifying the noun. The local ambiguity consists in the fact that the boundary between the main and relative clause may fall at any point between the left edge of the sentence and the immediately pre-verbal position (Japanese is a "super pro-drop" language, allowing both subjects and objects to be phonologically unexpressed under appropriate discourse conditions). We will see how this local ambiguity can be represented as a choice of lowering sites at the point where the parser receives the immediately post-clausal

[26]The abbreviations t and Øwill be used to stand for trace (of relativization) and pro (phonologically unexpressed pronominal element) respectively. We use subscripts to indicate case and coindexing. We believe that a "gapless" (Pickering and Barry, 1991) account of the Japanese data presented here would also be possible.

noun. In each of the three sentences in (4.9.a-c), the initial string may schematically be represented as follows:[27]

NP$_{nom}$ NP$_{acc}$ V$_{trans}$

On the assumption that the two NPs are initially structured as coarguments of the verb, the parser may have to "displace" one or more arguments from the clause on reaching the relativising noun. This is because a relative clause must contain a gap, which is coindexed with the head noun, and if the constituent to be relativised is already overt in the initially built clause, then that overt constituent will have to be displaced from the clause, and replaced with the gap. However, if the relativised constituent is not overt in the initially built clause, then no material will have to be displaced. For example, if the relativised constituent is represented by an empty *pro* category in the initially built clause, then the relativisation relation can be established by postulating the *pro* as the relativised trace, and coindexing it with the head-noun. Otherwise, if the relativised constituent is a non-overt adjunct in the initially built clause, then the relativisation relation can be established by adding the empty category representing the adjunct to the clause.

Inoue (1991) notes a general preference towards displacing the minimal amount of material from a completed clause to a higher clause. He calls this the "Minimal Expulsion Strategy".[28]

In the following, we will see how the "minimal expulsion" strategy can predict differences in the processing difficulty of the examples given in (4.9), discussed by Mazuka and Itoh (1995) and Inoue (1991).

The sentence in (4.9.a) is an example of a case where no material is displaced on reaching the head noun. This is because, although both arguments of the verb *kaihatusita* are present in the initial clause, the relativised constituent corresponds to a locative *adjunct* rather than an argument, so no overt constituents need to be displaced by postulating the relativised gap. Note that, in Japanese, relativisation of an adverbial is generally only possible if the role of that adverbial is temporal or locative. We will refer to such examples as "null-displacement". Null displacement examples of this kind reportedly do not cause conscious processing difficulty,

[27]We indicate an NP bearing case C as NP$_C$, and V$_{trans}$ denotes a transitive verb (i.e. a verb which takes one nominative and one accusative argument)

[28]We prefer to use the term "displacement" over "expulsion", since, as we shall see, in the present model, a displaced element is not expelled from the clause to which it is originally attached, but rather everything *except* the displaced element is lowered. However, we continue to use the term "minimal expulsion" when we refer to the processing strategy as proposed by Inoue.

according to (Inoue, 1991).

The sentence in (4.9.b) involves displacing *one* argument, i.e. the nominative NP *Yamasita ga*. At the point where the first verb is processed, a clausal structure will have been built, corresponding in meaning to "Yamasita visited his friend". However, on the subsequent input of the noun *siriai*, a gap has to be found in that clause. This time, since *siriai* "acquaintance" is not a plausible location or time for an action to take place, adverbial relativisation is not possible (consider the bizarre interpretation of the English NPs "the acquaintance *where* Yamasita visited his friend", and "the acquaintance *when* Yamasita visited his friend)". This means that the processor will be forced to postulate a gap in one of the two argument positions. Postulating the gap in the subject position causes the displacement of *Yamasita ga*, which results in the globally correct structure. We will call such examples "single displacement" sentences. According to Mazuka and Itoh (1995), sentences such as these so not cause conscious processing difficulty.

The sentence in (4.9.c) involves displacing *two* arguments, the subject argument and the object argument. We will postpone our discussion of how processing might proceed in such an example. Mazuka and Itoh (1995) claim that "double displacement" examples of the type exemplified in (4.9.c) *do* cause conscious processing difficulty.

There have been a number of proposals for explaining the relative difficulty of (4.9.c) in comparison to the other examples (Gorrell (1995b; 1995a), Weinberg (1993; 1995)).

In the following section, we will look at one of these proposals, due to Gorrell (1995b; 1995a), who proposes that the difference in difficulty between single and double expulsion is a direct consequence of his Structural Determinism. By contrast, we will argue that the relevant processing differences are more appropriately seen in terms of the preference strategy by which the parser implements reanalysis.

4.6.3 Gorrell's Explanation of Minimal Expulsion

Gorrell (1995b), (1995a) claims the contrast in processability between (4.9.b) and (4.9.c) can be derived via structural determinism (monotonicity of structural relations). However, we will see that, if we simply consider the structural relations before and after reanalysis, not only (4.9.b) but also (4.9.c) can be derived in a manner which preserves monotonicity.

Consider the schematic representation of a clause below:

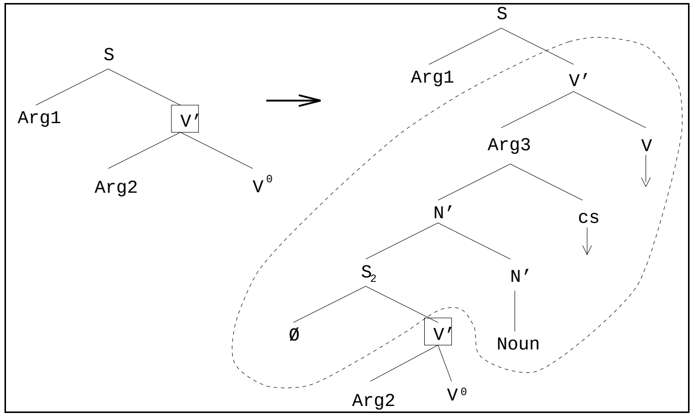
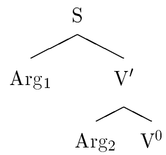


Figure 4.5: Single displacement via the application of lowering at V′ node. The material enclosed in the dotted line is inserted into the structure on reanalysis. The position marked by "cs" represents the case marker expected in the input.



Let us say that the next word in the input is a noun. If the argument corresponding to Arg_1 has to be relativised, then Arg_1 will have to be displaced by an empty argument (or trace). If the argument corresponding to Arg_2 has to be relativised, then both Arg_1 and Arg_2 have to be displaced. Let us consider the displacement of Arg_1 first. An alternative way of looking at this, as Gorrell (1995b), (1995a) has noted is that everything *except* Arg_1 is "lowered" (that is, in present terms, tree-lowering is performed on V′), and a new S node is created (call it S_2) which immediately dominates a newly created empty category in subject position. Then S_2 is adjoined as a premodifier to the noun. The noun is postulated to head Arg_3, which is attached as a coargument to Arg_1. This is illustrated in Fig. (5).

Now consider the displacement of both Arg_1 *and* Arg_2. Gorrell (1995b; 1995a) explains the difficulty of utterances requiring such double displacement in terms of the need to "delete"

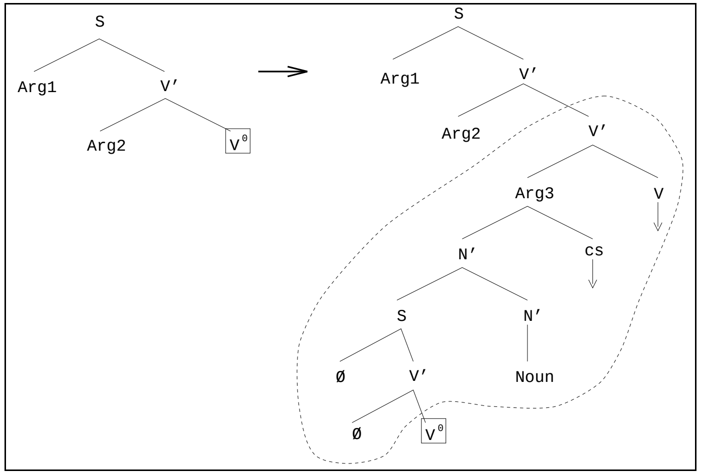


Figure 4.6: Double displacement via the application of lowering at V^0 node

a domination relation between (in our terms) V′ and Arg_2. In fact, however, this double displacement *can* be derived in an analogous manner to the single displacement example noted above. In this case, we lower the head node, V^0, and reconstruct a relative clause structure by adding the relevant nodes up to S_2, including *two* empty argument positions. This is illustrated in Fig. 6. As the reader can verify, this, as well as the single argument displacement preserves informational monotonicity, since the original position of V^0 dominates the post-reanalysis position.[29]

4.6.4 A Search-Based Explanation of Minimal Expulsion

In contrast to Gorrell's model (1995b; 1995a), where the distinction between single and double displacement is accounted for in terms of the parser's inability to withdraw structural

[29]In fact, the structure Gorrell proposes for this example includes an empty INFL node as a right sister of the VP. The double displacement example could be ruled out if we allow the verb to raise via verb-movement into this position. In this case, the last word to be processed will be in the INFL position, and thus, the V^0 node will no longer be accessible for lowering. However, this would rule out all reanalyses involving displacement of an object, but in some cases, such examples are possible without conscious processing difficulty, as we shall see in section 4.6.6. Furthermore, there is very little convincing evidence for the existence of an INFL projection in Japanese clausal syntax (Fukui (1986), see also Sells (1995)).

statements *at the initial point of disambiguation*, we would like to propose instead that the difficulty of examples such as (4.9.c) may be due to the parser initially performing a misreanalysis, which only becomes apparent at a later point of processing, from where recovery is difficult.

The standard definition for tree-lowering offered above (page 69) will obviously not suffice to deal with the reanalysis necessary in this type of example, since here we must add structure (including a new sentential node, and empty argument position(s)) which is not part of the subtree projection of the new head noun found in the input. The definition of tree-lowering has therefore been extended so that this extra structure can be built as part of the operation. The parser includes an *Argument Projection* operation, which, on the input of a head, checks for the presence of the required arguments, and, in the case of a verbal head, adds empty categories for any arguments which are missing, in cases where the grammar allows this. It is this operation which is employed in the extended definition of tree-lowering. Where lowering is applied to a head-projection, *Argument Projection* is reapplied, so that, in cases where the arguments of a verb are displaced by reanalysis, the embedded clause structure is "regrown", including any necessary empty argument positions (see Appendix). In the examples we have been considering, the "regrown" embedded clause can then be attached as a relative clause to the incoming noun[30], and this noun can then be attached as a coargument to the displaced arguments (Sturt (1994) contains details of this). Given the revised definition, either single or double displacement can be derived, depending on the node at which tree-lowering is applied. This means that, in order to account for the contrast in difficulty between (c) and (b), we will crucially have to appeal to the search strategy which the parser uses in finding a node for lowering. It will be clear that the bottom-up search we motivated in the previous section for English will predict exactly the opposite results for these Japanese examples. This is because the last word to be attached into a clause will be a verb, and therefore, at the point where the parser fails to attach the head noun, the lowest node accessible to the lowering operation will be the node immediately dominating this verb, i.e. in the above schemata, the V^0 node. This means that, if the parser begins its search of the accessible nodes at the bottom, the V^0 node will be the first to be tried, and (given that the original clause contains two arguments) the embedded clause will be reconstructed with two empty arguments, in effect, resulting in a

[30]In fact, it is *adjoined* rather than attached, since the relative clause *modifies* the head noun.

double displacement. On the other hand, single displacement, which is known to be easy for Japanese perceivers, will be predicted to be *more difficult*, since it corresponds to choosing a lowering site which is higher in the structure.

If we reconsider the single displacement example, (15.b), repeated below as (4.10), we see how the bottom-up search strategy wrongly predicts a conscious garden path effect:

(4.10) Yamasita ga yuuzin wo houmonsita siriai ni tegami wo kaita.
yamasita NOM friend ACC visited acquaintance DAT letter ACC wrote
"Yamasita wrote a letter to an acquaintance who visited his friend."

Taking the V^0 node immediately dominating *houmonsita* ("visited") as the node chosen for lowering, the parser will displace both subject (*Yamasita ga*) and object (*yuuzin wo*) into the main clause.

This will result in an ungrammatical continuation, in which the verb *kaita* ("wrote") takes two accusative arguments instead of one, and we must wrongly predict a garden path effect when the parser notices this downstream.[31]

(4.11) Yamasita ga yuuzin wo [$\emptyset_{nom}$ $t_{acc/i}$ houmonsita] siriai$_i$ ni tegami
yamasita NOM friend ACC visited acquaintance DAT letter
wo kaita.
ACC wrote

4.6.5 Top-Down Search

A moment's reflection reveals that, if we want to reproduce the "minimal expulsion" effects using the tree-lowering operation for the type of examples discussed here, we should define a preference to perform lowering at as high a site as possible. It therefore seems reasonable to postulate a top-down search for Japanese.

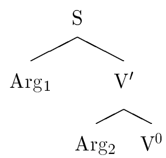
As yet we have very little on-line experimental data concerning the processing of this type of example,[32] and the following discussion is necessarily speculative, and should be seen as one possible way in which a top-down search could proceed. It should be noted in particular

[31] In fact this structure will violate the so-called *double-o constraint*, which bars the overt presence of two accusative marked NPs (in our terms, PPs) as arguments of the same predicate. (see Kuroda (1988) for details of this constraint).

[32] Though see footnote 35 below for a discussion of some very recent work.

that the parser may be sensitive to more types of information than the simply configurational issue of whether a node is high or low in a tree structure. One such factor may be the pragmatic plausibility of relativisation. Relativisation involves coindexing a head noun with an argument position (which is occupied by a gap site in the relative clause). Assuming a model which allows a certain degree of interaction of non-syntactic knowledge in making parsing decisions (c.f. Crain and Steedman (1985)), it may be that the processor takes into account the plausibility of establishing the referent of the head-noun in the argument position concerned. Other factors which may play a part here include the valency preferences of the verb, and the obliqueness of the argument to be relativised. Another factor which has been shown to be important is the case-marking on the relativising head-noun (Inoue (1991)).[33]

Imagine a two-argument clause has been built, resulting in the structure illustrated below, and that this is followed immediately by a noun, which must be incorporated somehow into the analysis.



One possible search strategy is that the parser considers each accessible node in a top-down order (S, V′, V^0), until a plausible relativisation site is found.

In a top-down search, the first node to be considered for lowering is S. This solution corresponds to retaining all arguments in the relative clause. We will call this *null-displacement*. Since the relative clause must contain a gap, and all of the arguments are overt, the gap must represent an adjunct, which, as we have seen above, must be either temporal or locative. Thus, the null-displacement option will only be available if the semantic content of the relativising head-noun is a plausible time or location for the semantic content of the relative clause to have taken place.

If this is not possible, we move on to consider node V′, and attempt to lower. This

[33] For example, native speakers seem to avoid on-line parsing decisions which would require postulating two nominative-marked NPs as arguments of the same predicate. This means that, in the class of examples which we have been discussing, if the relativising head-noun is marked with nominative case, then there will be a preference against displacing the (nominative marked) embedded subject to the matrix clause, i.e. a preference for the null-displacement option.

option corresponds to relativizing Arg_1. A new sentential node will be created, dominating an empty argument in the position occupied by Arg_1. This means that S will remain as the matrix sentential node, and will continue to immediately dominate Arg_1, or, to put it another way, Arg_1 will be displaced from the relative clause. This will be possible if it is plausible to coindex the referent of the head noun with the empty element in the position of Arg_1. If this is not possible, then the parser will descend to the next node, V^0, and attempt to relativise Arg_2. In the context shown above, this kind of search will predict the "minimal expulsion" strategy, with null-displacement preferred.

Consider first a null-displacement example, (4.9.a), repeated below as (4.12):

(4.12) [Mary ga sinseihin wo $t_{loc/i}$ kaihatusita] kaisya$_i$ ga tubureta.
Mary NOM new product ACC developed company NOM went bankrupt
"The company where Mary developed the new product went bankrupt."

At the point when *kaisya* (company) is found in the input, the first node to be considered will be the top S-node. The parser considers the relativisation corresponding to this node (i.e. adverbial relativization), which is found to be plausible, since a company is a plausible location for Mary to have developed a new product. Thus no difficulty is predicted, and indeed, this sentence does not cause conscious processing difficulty.

Now consider (4.9.b) repeated again as (4.13):

(4.13) Yamasita ga [$t_{nom/i}$ yuuzin wo houmonsita] siriai$_i$ ni tegami wo
yamasita NOM friend ACC visited acquaintance DAT letter ACC
kaita.
wrote

As before, the parser first builds the transitive clause with *houmonsita* ("visited") as the main verb. This time, null-displacement is not a possibility, since *siriai* ("acquaintance") is not a plausible location or time. This means that the processor considers the next node down as a lowering site. This node will be the constituent covering the object and verb *yuuzin wo houmonsita*. Accordingly, the subject argument, *yamasita ga* is displaced, and a relative clause structure is built with an empty subject position. This analysis remains grammatical throughout the parse, and the structure will be unproblematic for the processor.

Finally, consider (4.9.c) example repeated below as (4.14):

(4.14) Yamasita ga yuuzin wo [$\emptyset_{nom}$ $t_{acc/i}$ houmonsita] kaisya$_i$ de mikaketa.
Yamasita NOM friend ACC visited company LOC saw
"Yamasita saw his friend at the company he visited."

(4.14) is complicated by the fact that the string is not only locally but also globally ambiguous. The other reading is one in which the main clause contains two empty arguments, and the initially built clause remains intact as an adjunct relative. The null context strongly disfavours this reading, in which two uncontrolled gaps appear in the matrix clause, but it is possible to create a prior context which provides discourse control for both of these arguments, as in the question in (4.15.a) below. In this case, the utterance is considerably easier to process:

(4.15) a. anata wa doko no kaisya de Piitaa wo mikaketa no?
you TOP where GEN company LOC Peter ACC saw Q
"At which company did you see Peter?"

b. $\emptyset_{nom}$ $\emptyset_{acc}$ [$t_{loc/i}$ Yamasita ga yuuzin wo houmonsita] kaisya$_i$ de
Yamasita NOM friend ACC met company LOC saw
mikaketa.

"I saw him at the company where Yamasita visited his friend."

At the point where *houmonsita* ("visited") is attached, the parser will have built a simple transitive clause with both nominative and accusative arguments overt. On encountering the noun *kaisya*, the first option to be considered is the adjunct relativisation corresponding to null-displacement. This analysis is not implausible, since a company is a reasonable location for *Yamasita* to meet his friend. Let us say that the parser initially adopts this analysis. At the point when the final verb *mikaketa* ("saw") is encountered, neither of its nominative or accusative arguments is overtly present in the main clause. On the null context, this means that there are two uncontrolled arguments. However, raising the subject and object from the lower clause rectifies this situation. The two empty arguments in the lower clause are now both controlled, the accusative argument, marked $\emptyset_{acc}$ is grammatically controlled by the head noun of the relative, *kaisya*, and the nominative argument, $\emptyset_{nom}$, is pragmatically controlled by the matrix subject *Yamasita ga*. The explanation of the difficulty of the reading given in 4.14 is that this raising of the two arguments cannot be derived via tree-lowering at the verb region of the sentence. This is because, by the time the disambiguating final verb, *mikaketa* is encountered in the input, the relevant node for lowering will no longer be accessible, since it will be embedded inside the relative clause.

4.6.6 Easy Double Displacements

The present analysis predicts that double displacement should be possible if, at the initial point of reanalysis (i.e. where the immediately post-clausal noun is encountered), the need to perform lowering consistent with double displacement is obvious. This is in contrast to both Gorrell (1995b), (1995a) and Weinberg (1993; 1995), both of whom propose models which allow the displacement of an overt subject but not an overt object.[34]

Mazuka and Itoh (1995) give the following example, which reportedly causes no conscious processing difficulty, despite the fact that both the subject and object have to be displaced:

(4.16) Hirosi ga aidoru kasyu wo [$\emptyset_{nom}$ $t_{acc/i}$ kakusita] kamera$_i$ de totta.
Hirosi NOM popular singer ACC hid camera with photographed
"Hiroshi photographed the popular singer with the camera he was hiding."

We assume, as before, that the overt nominative and accusative arguments are initially structured as arguments of the verb *kakusita* ("hid"). On encountering the head noun *kamera*, the parser first considers the null-displacement option, which is found to be implausible ("the camera where/when Hirosi hid the popular singer"). The single displacement option is similarly ruled out ("the camera which hid the popular singer"). Finally, the double displacement option is considered, and is found to be plausible ("the camera which (somebody) hid"). This option is adopted, and the remaining processing proceeds without trouble.[35]

[34]Though as we have seen above, if the lowering of a verb is permitted, Gorrell's model *will* allow the displacement of an overt object.

[35] Data from a preliminary experimental investigation into the issues discussed in this section (Hirose, 1997) supports the general claim being made here, that the processor can easily perform double displacement if the need to do so is obvious at the point when the post-clausal head-noun is processed. Specifically, for sentences which eventually required double displacement, there was a significant reading time advantage in cases where the post-clausal noun could be only be relativized in a way which was pragmatically consistent with double displacement, as compared to cases which were initially ambiguous between single and double displacement. However, the data from this experiment do not appear to support the simple serial search mechanism proposed here, since initially ambiguous *single displacement* examples were processed with a difficulty equal to that of the ambiguous double displacements. However, these data should not be seen as conclusive, since no study has manipulated ambiguity and reanalysis type in a fully controlled 2×2 design. We leave the resolution of this issue for further research.

A similar effect can be seen if we consider topicalization. In Japanese, topicalised elements, which are given an overt morphological marker *wa*, almost invariably occur in the matrix clause[36] though they may control a "gap" at any level of embedding. Below we reproduce the double displacement example (4.9.c) with the nominative marked argument topicalised. This is reported to be considerably easier to process than the non-topicalised version.

(4.17) Yamasita wa yuuzin wo [$\emptyset_{nom}$ $t_{acc/i}$ houmonsita] kaisya$_i$ de mikaketa.
Yamasita TOP friend ACC visited company LOC saw
"(as for) Yamasita (he) saw his friend at the company he visited."

In the top-down search described above, we hypothesised that, on reaching the head noun, *kaisya* ("company") the processor first attempts to form a relative clause consistent with null-displacement, with a locative relativisation reading. However, in (4.17), the parser can immediately eliminate this option, since it would involve a topicalised phrase *Yamasita wa* appearing in a subordinate clause. The next option to be tried will be the *single displacement* option, in which the constituent covering *yuuzin wo houmonsita* ("visited the friend") is lowered. However, this may be discounted on the grounds of plausibility, since "company" is not a plausible subject for "visited". We assume that the parser then adopts the double displacement option, which eventually turns out to be correct. Thus (4.17) is correctly predicted to be easier than its non-topicalised counterpart, (4.9.c). However, on the bottom-up search, there would be no difference predicted, since the V^0 node will be the first node chosen as a prospective lowering site in both cases.[37]

[36]Though see Kuroda (1988), for some limited exceptions.

[37]Robert Frank (p.c.) points out that (4.17) can actually be processed with a single-displacement reanalysis. In this case, *kaisya* is initially taken to be modified by a relative clause containting the accusative NP and the verb, and *kaisya* is interpreted as a locative within this clause ("the company where he visited his friend"). If the parser initially adopts this (locative) single displacement option as its preferred alternative, then we need to find an alternative explanation for the ease of processing (4.17) in comparison with the non-topicalised alternative (4.9.c), since in both cases, the parser will have to perform non-monotonic reanalysis at the final verb in order to obtain the correct reading. One possible explanation for the difference might be that although both sentences will require a non-monotonic reanalysis at the final verb, this reanalysis will involve breaking *two* dependencies in the case of the non-topicalized version (4.9.c), but only *one* dependency in the case of the topicalized version (4.17), since the initial processing decision at the head noun *kaisya* will have involved null-displacement in the case of (4.9.c), but single displacement in the case of (4.17).

4.7 Explaining Differences in Search Strategies

In the preceding sections we have saw how the currently known data motivate a difference in search strategy for reanalysis in English and Japanese. In this section, we look in detail at the effect which reanalysis has on linguistic dependencies, and suggest that, rather than being explicitly parameterised in the parser, the difference in search strategies may be motivated by a strategy on the part of the parser to preserve as many dependencies as possible. It will be noticed that this account shares features with the Minimal Revisions Principle (Frazier, 1990a; Frazier, 1994) discussed in chapter 2, page 31.

4.7.1 Three Types of Lowering

As discussed above, the use of informational monotonicity at the level of purely structural relations may result in non-monotonic behaviour at the level of "secondary" linguistic dependencies. In this section we discuss three different scenarios in which lowering may apply, and in each case, we discuss the consequences for the dependencies between head and satellites.[38] Note that the lowering operation discussed in this chapter is what we may call *retrospective* lowering; that is where the disambiguating word is *preceded* by the constituent that it lowers. Another possible operation, which we do not discuss here is *anticipatory* lowering. In this case, the disambiguating word will precede the lowered constituent, where all the terminal nodes dominated by the lowered constituent are dangling nodes not yet dominating lexical material (i.e. an adjunction-like operation is performed on a predicted branch of the tree, as yet not dominating any lexical input).[39] The parser described in chapter 6 allows such "anticipatory lowering". These three types of lowering, namely *head extension*, *satellite detachment* and *head detachment*, are illustrated in figures 4.7, 4.8 and 4.9 respectively.

4.7.2 Head-Final and Head-Initial Languages

Now let us consider the consequences of performing each of the three lowering operations in head-initial and head-final languages. We consider in particular the number of dependencies which must be broken in order to perform lowering on the above structures. By "depen-

[38]By "satellite", we mean any complement or adjunct which depends on the head.

[39]This would be required in an incremental parse of a sentence such as *John likes tasty apples.*, where the adjective *tasty* is adjoined to a noun projection which is not yet justified by lexical input.

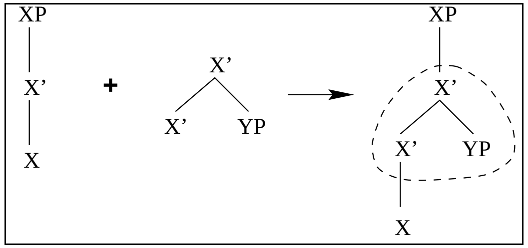


Figure 4.7: HEAD EXTENSION: A head projection is extended through the insertion of material at an intermediate point. Example: post-modifier attachment.

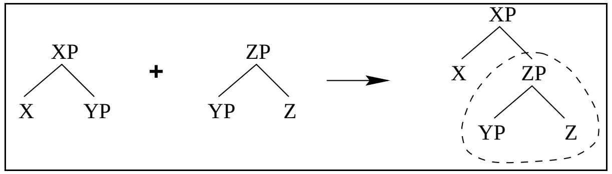


Figure 4.8: SATELLITE DETACHMENT: A satellite projection is broken. Example: *John knows the truth hurts*, where the satellite projection between the NP dominating *the truth* and the matrix VP is broken.

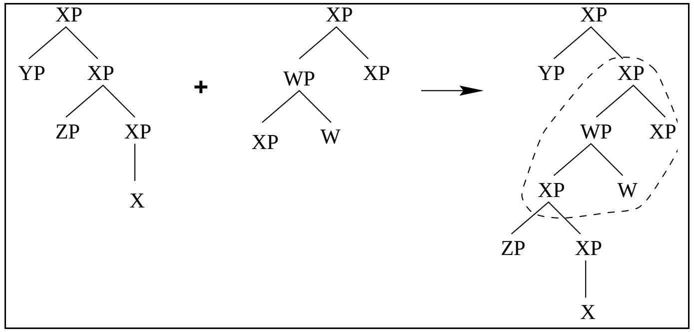


Figure 4.9: HEAD DETACHMENT: A head projection is split into two separate projections. Examples in discussion of Japanese data.

dency", we mean a (licensing) relation between a satellite and its head *both of which have been encountered in the input so far*. In other words, we do not consider a dependency to exist between a head and its satellite if either of the two have not yet been encountered in the input.[40]

Let us consider *head extension* (figure 4.7) first. On standard X-bar assumptions, this corresponds to the attachment of a modifier. In a head-final language, in which only pre-modifiers, and no post-modifiers exist, this operation can only be employed if the word which heads the projection to be extended has not yet been encountered in the input, and not via retrospective lowering as we have defined it. In a head-initial, or mixed language which allows post-modifiers, like English, however, head-extension will be possible via retrospective lowering. In this case the lowering operation will *add* one dependency between the head and the newly adjoined constituent.

Now consider *satellite detachment* (figure 4.8). In a head-final language, all satellites precede the head. This means that, once the head has been attached, none of the satellites will be *accessible* in the sense defined above (page 68), since none will dominate the head (i.e. the last word to be incorporated). So in a head final language, satellite detachment will only be possible in cases where the licensing head has not yet been reached in the input, and therefore will not break a dependency; that is to say that satellite detachment will be restricted to cases of *anticipatory lowering*. In a head-initial language, by contrast, any satellite which is preceded by its head will remain accessible to immediately following lexical material when the satellite phrase in question has been completed, and subsequent detachment of the satellite will result in breaking the dependency between that satellite and its head.

Finally, consider *Head detachment* (figure 4.9). This is very likely to be found in a head-final language. It corresponds to the case where a constituent, call it `XP`, has been completed, but the word subsequently found in the input (call it `W`), requires one of the nodes on `X`'s head projection. This node is attached to `W`'s left, and replaced with the root node of `W`'s subtree projection. The word `W`, requiring a constituent on its left, may be a postposition, for example, in a head-final language. Head detachment may break any number of dependencies. In a head-final language, at the point where a constituent has been built with a head and all

[40]Note that, in modular theories such as GB, there may be a number of licensing relations between a single head and a single satellite. In this case we do not count each single licensing relation as a single dependency, but treat the entire complex of licensing relations as one dependency.

its satellites, but no further attachments have been made, all nodes on the head projection of that constituent will be accessible. If a head-detaching lowering operation subsequently has to be performed, then we will assume that the processor attempts to lower at a node consistent with breaking the smallest number of dependencies possible, and this will coincide with Inoue's "minimal expulsion strategy".

It can be seen that the minimal expulsion preference can be derived via a preference to lower at the highest node possible, thus maintaining intact the largest number of dependencies between the head and its satellites. The reanalysis of an initially built clause as a relative in the examples we have been discussing above may be seen as instances of the schema in 3. in section 4.7.1 above (abstracting away from syntactic details), where the word `W` corresponds to the post-clausal head noun, and `XP` is the clausal (verbal) projection.

In English, on the other hand, head extension and satellite detachment will be employed far more often. Since, as far as the number of broken dependencies is concerned, nothing hinges on the choice of lowering site for either of these, the processor may be following a different strategy, which may include a preference to lower nodes which have been created as recently as possible (i.e. assuming a right-branching structure, to choose a low site). Considerations such as these may well underlie the differing search strategies between the two languages.

4.8 Other Computational Approaches to Human Syntactic Reanalysis

Suzanne Stevenson's competition based model (1993; 1994a) derives from the constraints of a connectionist architecture a number of very interesting predictions for syntactic processing. The model is similar to that proposed here in the sense that reanalysis is not seen as qualitatively different from simple attachment. In fact Stevenson's constraint that reanalysis should only involve nodes on the right edge of the tree, which follows from the space constraints inherent in the connectionist architecture, leads to a reanalysis operation which is very similar to tree-lowering constrained by *accessibility*, as defined here. The decay of network activation predicts an empirically supported recency preference in reanalysis as well as attachment, which means that the "bottom up" search strategy which we have discussed here does not have to be stipulated. However, it is not clear how the model would perform in processing head-final languages, where, assuming an incremental processing regime, the recency preference would presumably predict the opposite of the "top-down", dependency preserving search strategy motivated for Japanese in this chapter. That is to say that a strategy which takes into account more grammatical information may be required in addition to the simple recency preference which is predicted in a decay of activation model. Also, the constraints of Stevenson's model may in fact be too strong. For example, phrase structure cannot be postulated without explicit evidence from the input, so that, for example, in sentence such as *Mary thinks John likes oranges*, processing difficulty is predicted on the input of *John*, because *thinks* does not select for an NP, and therefore no attachment site can be postulated from explicit evidence in the input up to that point in processing. On the other hand, in a model such as we have been discussing in this chapter, which does not forbid non-lexically driven prediction, this problem need not arise, as we have seen. This strict constraint also makes it difficult to extend Stevenson's model to include more complex reanalysis operations, such as the extended version of Tree Lowering described in section 4.6 of this chapter.

Richard Lewis (1993) presents a comprehension model, NL-SOAR, which incorporates a syntactic reanalysis component. If, on the attachment of an incoming word, an inconsistency is detected within the local maximal projection to which the incoming word is attached, then NL-SOAR's "snip" operator can break a previous attachment within this maximal projection and reattach it elsewhere in the tree. This operation is more powerful than Tree-Lowering in the sense that the phrase detached by the snip operator does not have to be reattached in a position which is dominated by the projection of the incoming, disambiguating word. This results in an impressive range of correct predictions. For example, in a sentence such as *Is the block on the table red?*, the disambiguating word *red* can trigger the snip operator to detach the PP *on the table*, which has previously been attached as the complement of the copula, and reattach it as an adjunct of *block*, correctly predicting this sentence to cause no processing difficulty, while lowering will not account for this, since the post-reanalysis position of *on the table* is not dominated by the projection of the disambiguating word *red*. However, NL-SOAR does overgenerate[41] on a class of examples such as: *"The psychologist told the woman that he*

[41] By 'overgenerate' here, we mean that the parser is able to process examples which are actually conscious garden paths.

was having trouble with to leave." and *"The boy put the book on the table on the shelf"*, which both involve reattaching material into a preceding phrase, which is not possible in Gorrell's model, and also cannot be generated via tree-lowering.

4.9 Conclusions

This chapter has aimed to show how we can retain the intuitive appeal of a D-theory based approach, while also providing an explanation of the limitations on the human parser's ability to reanalyse structure in terms of a constrained search space.

One of the hall-marks of the D-theory based approach is that there is no major conceptual distinction between (unconscious) reanalysis and attachment. The definition of a *reanalysis* oriented attachment operation (i.e. *tree-lowering*) has led us to explore the possibility that there are "reanalysis ambiguities" just as there are attachment ambiguities, and this demands the consideration of preferences for the application of the tree-lowering operation, which have to be defined over and above the basic D-theoretic machinery. We have seen that the preference to "reanalyse low" in English has to be replaced by a preference to "reanalyse high" in Japanese, and we have speculated that this may be due to a "dependency preserving" strategy of the parser.

Appendix

The purpose of this appendix is to give a more precise characterisation of the implementation by giving a brief description of the basic algorithm and data structures employed.

Grammar Representation

Each word in the lexicon is associated with an argument frame, and a lexical category (e.g. verb, noun). To each lexical category corresponds a *grammar entry*, which is a quintuple, $\langle C, D, R, Left, Right \rangle$, where C is the name of the lexical category (e.g. V^0, N^0, etc), D is the set of dominance and precedence relations which describe the lexical subtree anchored in C, R is the root node of the lexical subtree, and *Left* and *Right* are lists of the nodes in the subtree which are attachment sites for the category C, where the nodes in *Left* precede C and the nodes in *Right* follow C.

The grammar entry for the verb category in English, mentioned above in section 4.3.1, is as follows:

$C = \mathrm{V}^0$

$D = \{\mathrm{dom(VP,V^0),dom(V^0,Word),dom(S,VP),dom(S,NP),prec(NP,VP)}\}$

$R = \mathrm{S}$

$Left = \langle \mathrm{NP} \rangle$

$Right = \langle\ \rangle$

Note that the grammar entry only represents properties of the general lexical category in question. Thus, the grammar entry for verbs, shown above, has an attachment site for a subject, reflecting the fact that all verbs in English must have a subject, but does not include attachment sites for the internal arguments of particular verbs. These attachment sites are projected with reference to the argument frames of each particular verb in the lexicon, when the verb is used.

Global State

The global state of the parser is a triple, $\langle D, Root, Right \rangle$, where D is the set of dominance and precedence relations which describe the global tree built up so far, *Root* is the root node of the global tree built up so far, and *Right* is the list of unsaturated attachment sites, which will have to be satisfied via *Right Attachment*. (Note that, by definition, there cannot be unsaturated left attachment sites in the global tree, since these will be inaccessible according to the theory of trees.)

The attachment of a new word is achieved by equating the relevant nodes, as required by the definitions of Attachment and Tree-Lowering, adding the new structural relations from the word's grammar entry, projecting any necessary obligatory arguments, and, if necessary, updating the global state so that it represents the relevant new right attachment sites, and new root node.

Argument Projection

When the parser incorporates a new word into the description, the global state is updated with reference to the new word's argument frame. Recall the grammar entry for English verbs, given above. If an obligatorily transitive verb is processed, the resultant global state

will include a new NP node (call it NP_2) in the list of right attachment sites. Also, the following relations will be added to the global description to encode a "dangling node".

$\{\mathrm{dom(VP,NP_2),\ prec(V^0,NP_2)}\}$

In the case of Japanese, which allows empty arguments, any arguments "missing" from the left of the new word will be added as empty arguments, and, if necessary, the root node will be updated. For example, if a single accusative NP is followed by a ditransitive verb in the input, then the "missing" dative and nominative arguments will be added to the global description as empty categories, and the verbal projection extended to the S node dominating the subject NP.

Non-lexical Structure Building

The grammar representation used here is not entirely lexicalised; that is to say that there are grammatical objects manipulated by the parser which do not have a lexical anchor. This is employed to deal with Japanese relative clauses, which have no overt relative pronouns, or other explicit lexical signals to show that they can combine with following nouns. In the current implementation, we allow a clause containing at least one empty category to be "extended" to become a noun modifier, (basically, the clause $[_S \ldots]$ is converted to $[_{N'}[_S \ldots]\ [_{N'}]]$ by adding the appropriate relations, and updating the root category as appropriate. A similar strategy would be needed to deal with English reduced relatives, for example.

As mentioned in 4.4, we also allow a projected clausal complement in English to be extended down to the subject NP of the complement clause , even when the input lacks an explicit complementizer. This is also a case of building structure not directly justified by lexical input.

Basic Algorithm

The following is a pseudocode description of the basic control structure of the algorithm.

1. • **If** Input is empty,
   • **Then** succeed.
   • **Else** go to 2.
2. **Project Word**

(a) Read next word W.

(b) Find category C of W.

(c) Find grammar entry GE for C.

(d) Go to 3.

3. **Attachment**

(a)
- **If** the Global State is undefined (i.e. W is the first word of the input), and W does not require overt arguments to its left,
- **Then** instantiate GE as the new Global Description, apply *Argument Projection*, and go to 2.
- **Else**, go to 3b.

(b)
- **If** the preconditions for *Right Attachment* are met,
- **Then** combine GE with the current Global Description via *Right Attachment*, perform *Argument Projection*, and go to 2.
- **Else** go to 3c.

(c)
- **If** the preconditions for *Left Attachment* are met,
- **Then** combine GE with the current Global Description via *Left Attachment*, perform *Argument Projection*, and go to 2.
- **Else**, go to 4.

4. **Tree Lowering**

(a) Find the *Current Node Path* P, i.e. the set of accessible nodes, ordered according to the ordering function F (see below for explanation).

(b)
- **If** P is empty,
- **Then** fail.
- **Else**, go to 4c.

(c) Remove N, first node of P, leaving $Rest$, the remainder of P.

- **If** the preconditions for *Tree Lowering* are met at N,
- **Then** apply *Tree Lowering* at N, and perform *Argument Projection*.

- **Else**, set $Rest = P$, and go to 4b.

The different search strategies for the application of Tree Lowering in English and Japanese are captured by setting different values to the ordering function F, mentioned in 4a. To obtain the bottom-up search used for English, F orders the relevant nodes from bottom to top, while the top-down search used for Japanese orders them from top to bottom. Though this difference in search strategies is presented as a stipulation in the actual implementation, we strongly suspect that it is actually the result of some dependency preserving strategy, as outlined in section 4.7.1 above.

It should also be noted that the algorithm, as presented above, is slightly simplified. In fact, the current implementation allows the Global Tree Description to be extended via non-lexical structure building, as briefly described in the previous subsection, before applying the composition operations. Furthermore, in cases where Tree-Lowering is applied on a head projection node, (i.e. where *Head Detachment* is applied, (section 4.7.1)), *Argument Projection* is re-applied to the head word of the detached projection, in order to "regrow" empty arguments that will replace those "displaced" by the operation. This, as well as non-lexical structure building, allows us to capture the rather complex version of Tree Lowering described in section 4.6.

It should be noted that the serial character of the algorithm results in a preference for Right over Left Attachment (since Right Attachment is tried first). We do not have any strong commitment to this preference, and we have not found any phenomena which exhibit a Right/Left Attachment ambiguity, for which the preference would make crucial predictions.[42]

[42] Abney (1989) reports similarly that he finds no phenomena which exhibit the Attach/Attach-L conflict in his Licensing based model.

Chapter 5

Monotonicity and Thematic Structure

5.1 Introduction

The model which we described in the previous chapter represents a particularly restrictive instantiation of the monotonicity hypothesis. In this chapter,[1] we will identify data which demonstrates that the model is *too* restrictive, thus requiring us to loosen the constraints on which the model is built. We will argue that, rather than rejecting the monotonicity hypothesis, the problems can be solved by defining the monotonicity constraints not over purely *constituent structure* representations, as previous models, including that introduced in the previous chapter have done, but over *thematic* representations. This allows us to incorporate insights from Pritchett (1988) and Frazier and Clifton (1996) into the framework.

Before we proceed, we should point out that this chapter does not attempt to provide an *algorithmic* level theory, in the sense of Marr's (1982) levels of explanation for theories of mental processes. Rather, it is intended as a *computational* level theory, which specifies constraints of reanalysis in an implementation-independent manner. In this sense, the model described here is couched at a level of description that is similar to Gorrell's (1995b) model that we discussed in chapters 3 and 4.

Of course, as we have argued above, a full model of reanalysis needs to include a specification of the search strategies through which revision takes place, and it is intended that the *computational level* model described here will in the future be supplemented by an *algorithmic level* implementation, in the same way that the implementation described in chapter 4 can be seen as an *algorithmic level* extension of Gorrell's *computational level* theory.

[1]The material in this chapter is largely drawn from Sturt and Crocker (1997).

In chapter 6, we will discuss some initial work on a parser which is intended to serve as a basis for the implementation of the model that we will describe in this chapter. The reader will notice that the changes which we discuss in this chapter will result in a model which is less restrictive than that reported in chapter 4, in the sense that a wider range of reanalyses will be made possible. This means that, an *algorithmic level* implementation of the model will have to allow a more powerful set of parsing operations than that the set described in chapter 4.

5.2 Modifier Attachment

Since Cuetos and Mitchell's (1988) study, a great deal of research has been directed at questions concerning modifier attachment, and the results of this research have had far-reaching consequences for theories of sentence comprehension in general. Within the Garden Path tradition (Frazier, 1978) this has led to a principled distinction between the way in which *modifiers* are treated, and the way in which *obligatory constituents* are treated (Frazier, 1990b), leading to the distinction between primary and secondary phrases in Construal Theory (Frazier and Clifton, 1996).[2] As we point out in this section, results from recent studies on modifier attachment pose problems for standard monotonic approaches as well. In sections 5.5 and 5.3.1, we will show how the Construal notion of Primary and Secondary phrases can be integrated with a monotonic model to solve these problems.

Consider the following well-known sentence from Cuetos and Mitchell (1988).

(5.1) The journalist interviewed the daughter of the colonel who had had the accident.

[2]Hemforth, Konieczny, and Scheepers (1995) have argued that the subclass of *relative clauses*, rather than the wider class of modifiers in general, are treated in a distinguished manner by the processor, as the attachment of relative clauses involves anaphoric, as well as syntactic processes. While we take seriously the claim that *some* types of constituent are to be treated specially, and attempt to define the nature of this special treatment, we will not discuss in any great detail the question of whether this subclass should be equivalent with modifiers in general, or of relative clauses in particular, though we point out here that recent data (Traxler, Pickering, and Clifton, 1996) are compatible with Hemforth et al's claim.

There are two possible interpretations for sentences such as (5.1), corresponding to the attachment of the relative clause as a modifier of *daughter* on the one hand (high attachment) and as a modifier of *colonel* on the other (low attachment). Notice that, if the relative clause is initially attached in the low site, and subsequently reanalysed to the high site, then, at the level of constituent structure, this reanalysis will break a dominance relation. If the relative clause is attached in the low site, then it will be dominated by the NP headed by *colonel* (we will call this NP3). However, reanalysis to the high site will remove the relative clause to a position where it is no longer dominated by NP3.

Thus, standard, constituent structure based D-theory models make a clear prediction that reanalysis of this type should cause conscious processing difficulty.

However, recent experimental evidence from Italian (De Vincenzi and Job, 1993; De Vincenzi and Job, 1995) and Japanese (Kamide and Mitchell, 1996), suggests that, in sentences similar to 5.1 in these languages, although the initial, on-line preference is for for the *low* site, the final, off-line preference is for the *high* site. Specifically, questionnaire studies in both Italian (De Vincenzi and Job, 1993) and Japanese (Kamide and Mitchell, 1996) show a high attachment preference for globally ambiguous sentences similar to (5.1). However, self-paced reading studies in both languages (De Vincenzi and Job, 1993; De Vincenzi and Job, 1995; Kamide and Mitchell, 1996) showed evidence for an initial on-line commitment to the low attachment site, where the sentences are disambiguated syntactically in the case of De Vincenzi and Job, and pragmatically in the case of Kamide and Mitchell. In De Vincenzi and Job's self paced reading studies, comprehension questions were also added, and accuracy was found to be greater for high attached sentences than low attached sentences (De Vincenzi and Job, 1995), further illustrating the discrepancy between the initial and the final preference.

This suggests that, at least in the languages studied by these authors, the revision from low attachment to high attachment, far from causing conscious processing difficulty, as is predicted by standard monotonic models, is actually the preferred course of action for the processor. We will discuss this phenomenon in section 5.3.1.

5.3 Thematic Representation

We believe that the problem outlined in the previous section is a result of defining the D-theory constraints over exclusively constituent structure representations, so that all nodes in

a tree are treated equally. It may be possible, therefore, to solve some of these problems by defining the D-theory constraints over different representations.

A number of researchers have suggested that thematic structure plays a crucial role in language comprehension (Abney, 1989; Crocker, 1996; Frazier and Clifton, 1996; Pritchett, 1988; Pritchett, 1992; Gibson, 1991).

Assuming that this suggestion is correct, then we should expect a representation-preserving language processor to be particularly reluctant to alter thematic structure during reanalysis. In this section, we will look at the Construal Principle as defined by Frazier and Clifton (1996). We will also suggest some similarities between these constraints and monotonic models, which will be developed in subsequent sections.

5.3.1 Construal: Thematic Assignment and Modifiers

The following is the "Construal Principle" from Frazier and Clifton (1996)

Construal Principle:

i. Associate XP, where XP cannot be analysed as instantiating a primary relation, into the current thematic processing domain.

ii. Interpret XP within that domain using structural and non-structural (interpretive) principles.

Current thematic processing domain
The current thematic processing domain is the extended maximal projection of the last theta assigner.

Primary phrases and relations include

a. the subject and main predicate of any (+ or -) finite clause.

b. complements and obligatory constituents of primary phrases.

The Construal Principle concerns the initial incorporation of non-primary phrases into the representation, and is not explicitly stated as a constraint on reanalysis. However, under standard assumptions, the principle predicts a processing cost associated with "removing" a non-primary phrase from its "current thematic processing domain". We illustrate this in figure 5.1.

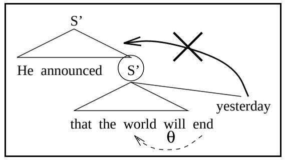


Figure 5.1: Informal illustration of the constraint on reanalysis implied by Frazier and Clifton's *Construal Principle*. The circle surrounding the S′ indicates that this node is the extended maximal projection of the theta assigner *end*.

Frazier and Clifton's Construal Principle, predicts a cost in cases where a constituent C has to be moved from outside a thematic domain D. This proposal seems to share a great deal in common with monotonic models; in D-theory terms, we can say that Construal predicts a cost for breaking a dominance relation between the extended maximal projection of a thematic role assigner and a constituent C.

Now, the insight behind the Construal Principle gives us an opportunity to address the modifier attachment problem that we mentioned above. This is hardly surprising, since it was with precisely these phenomena in mind that the Construal Principle was originally defined. However, in subsequent sections, we will go on to show how the insight can be incorporated into a monotonic model, and, moreover, we will demonstrate that the monotonic approach has empirical and descriptive advantages over the Construal Principle in its coverage of modifier attachment in left-branching constructions. Consider figure 5.2, which shows the English translation[3] of an Italian attachment ambiguity from De Vincenzi and Job (1995). Many authors assume a notion of thematic structure where, in a complex NP of the type illustrated in Figure 5.2, a thematic role is assigned from the higher noun to the lower NP (Gibson, 1991; De Vincenzi and Job, 1995; Gilboy et al., 1995; Frazier and Clifton, 1996). That is, in the case of the sentence illustrated in Figure 5.2, the noun *father* assigns a role to the NP headed by *girl*. The preposition *of* does not assign a thematic role of its own, but appears for purely configurational reasons (perhaps to assign case to the lower NP, or as a "reflex" of the lower NP's inherent case (Chomsky, 1986b)). If this is the case, then it can be readily ascertained that, at the point where the relative clause is encountered, the current

[3]The English translation is made purely for convenience, and not to suggest that the initial low attachment preference also exists in English.

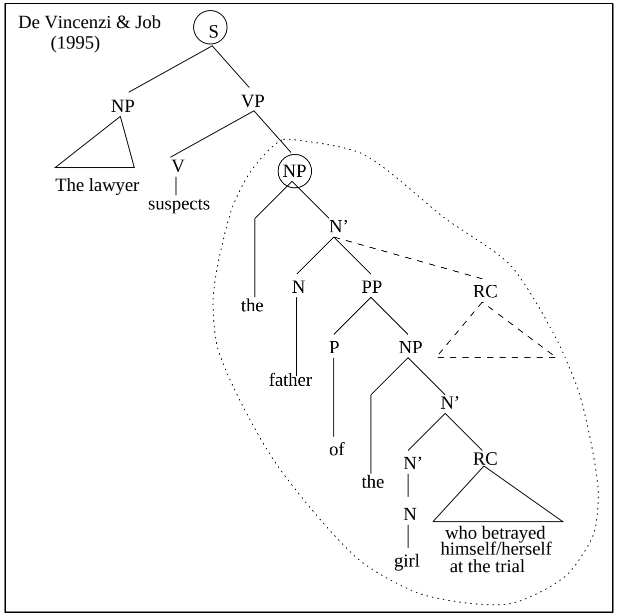


Figure 5.2: A relative clause attachment ambiguity. The dotted line encloses the locus of possible re-attachment sites for the relative clause once it has been attached in the low site. The high site is available here.

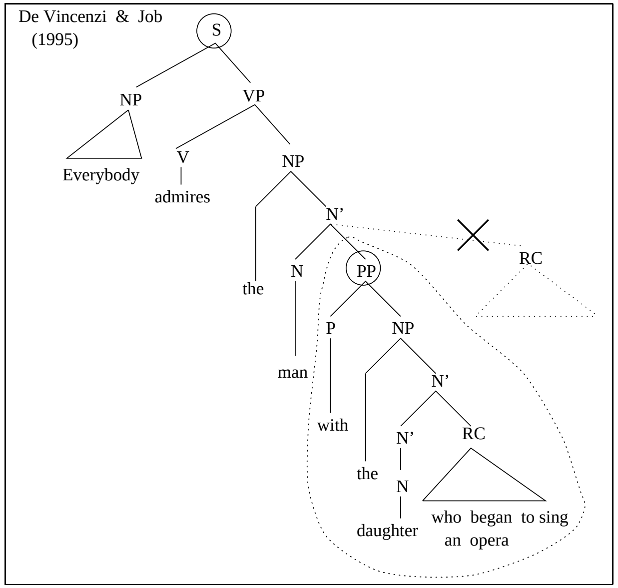


Figure 5.3: An illustration of a relative clause attachment ambiguity involving a preposition which assigns a thematic role. The locus of possible attachment sites for the relative clause now excludes the high site

thematic processing domain in this configuration consists of the subtree which has the circled NP as its root, since this is the maximal projection of the theta assigning head *father*. The consequence for this is that the thematic processing domain includes both the high and the low attachment sites for the relative clause. Therefore, Construal predicts that no processing cost will be associated with interpreting the relative clause as attached in either of these two positions. In contrast, consider the situation in which the preposition inside the complex NP assigns a thematic role in its own right, shown in figure 5.3. In this case, it is claimed that because the preposition *with* assigns a role to the lower NP, the maximal projection of *with* (i.e. PP) is a thematic processing domain. Construal therefore predicts that, once the relative clause has been attached in the low site, it cannot be re-attached to the high site without causing measurable processing difficulty. This pattern is consistent with De Vincenzi

and Job's finding that, although an initial low attachment was found for the relative clause, for conditions including both non-thematic and thematic prepositions (illustrated by figures 5.2 and 5.3 respectively), a final, off-line low preference was found only for the condition including the *thematic* preposition (illustrated by 5.3). Similar effects of the thematic status of the preposition (or postposition) on the attachment behaviour of modifiers in complex NPs has been found in a number of different studies, in English (Traxler, Pickering, and Clifton (1996)) Japanese (Branigan, Sturt, and Matsumoto-Sturt (1996), see chapter 8 of this thesis) and French (Baccino, De Vincenzi, and Job, 1996).

At this point, it is worth pointing out a subtle distinction between the effect found by De Vincenzi and Job (1995) and the predictions of Construal Theory. Taken together, the data of de Vincenzi and Job (1993; 1995) suggest that Italian speakers employ a *recency* strategy in their initial attachment of relative clauses, but that the processor subsequently *prefers* to reanalyse to a high attachment of the relative clause, just as long as the higher site is within the same minimal thematic domain as the original attachment site. In contrast, the Construal Principle defines a form of *underspecified* attachment, in which all modifier attachment sites within the current thematic processing domain are equally preferred from a *syntactic point* of view, and ambiguities are resolved with reference to higher level, semantic and pragmatic principles. This means that the initial recency preference found by de Vincenzi and Job in their on-line experiments is not accounted for in Construal theory. However, contrasting with De Vincenzi and Job's findings, recent eye-tracking experiments in English (Traxler, Pickering, and Clifton, 1996) have found evidence which is more strongly supportive of Construal's underspecified attachment strategy. Specifically, in sentences where the high and low attachment site appears in the same Thematic Processing Domain, the authors found a processing advantage for globally ambiguous sentences in comparison with both high and low disambiguated sentences. By contrast, in a further study, in which the complex NP included the preposition *with*, (i.e. only the low attachment site appeared in the current Thematic Processing Domain), the low attachment and unambiguous conditions were processed more easily than the high attachment condition. Taken together these results are compatible with a strategy in which multiple analyses are considered in a non-competitive fashion,[4] but these analyses are limited to attachments within the current thematic processing domain. This

[4]Competition would predict increased cost for the ambiguous condition.

should be contrasted with the findings of De Vincenzi and Job (1995), which, as we have seen above, appear to suggest that, initially, only the low attachment analysis is available, even in cases where an alternative high attachment site is available within the same thematic processing domain.

At present it is impossible to say whether or not the discrepancy between de Vincenzi and Job's results on the one hand, and Traxler et al's results on the other reflects a genuine processing difference (for example, a cross linguistic difference in processing strategies). However, we will assume that, at least in some cases, a processing theory needs to be able to account for cases in which the attachment of a relative clause is preferentially revised from a low site to a high site, and that preference to perform such a revision is affected by thematic structure.

5.3.2 Generalizing Monotonicity in Terms of Properties

Recall Gorrell's (1995b) *Structural Determinism* constraint. This predicts that processing cost is associated with reanalysis which deletes either precedence or dominance relations. In the following, we again use the abbreviation "$S \Longrightarrow T$" to stand for the transition from state S to state T.

Dominance Constraint:

For each transition $S \Longrightarrow T$, if X dominates Y at state S, then X dominates Y at state T.

Precedence Constraint:

For each transition $S \Longrightarrow T$, if X precedes Y at state S, then X precedes Y at state T.

Since the dominance and precedence relations constrained by this definition of monotonicity are between all nodes, this can be seen as a definition of monotonicity at the *constituent structure* level of representation.

In this chapter, we will present two methods of generalizing monotonicity, which will allow us to define the constraints in terms of other representations. The inspiration for both of these methods comes from Barker and Pullum (1990), who introduce two similar methods of generalizing *command relations*, based on *properties* on the one hand, and *relations* on the other. In this section, we describe a definition of monotonicity which is generalized in terms

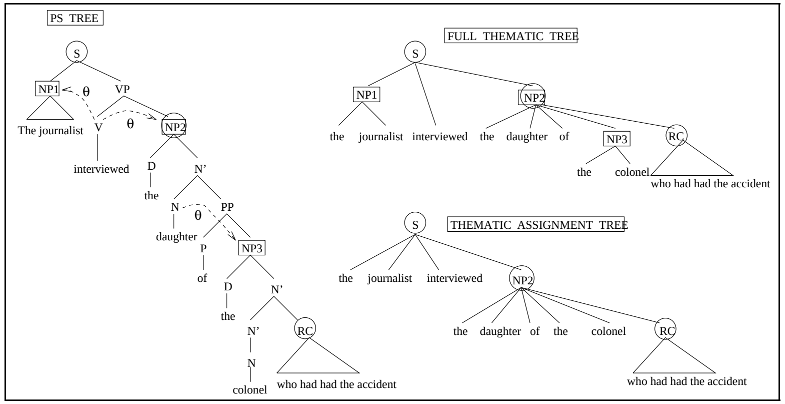


Figure 5.4: A representation showing thematic *assigner* nodes (enclosed in circles) and thematic *receiver* nodes (enclosed in squares) for a modifier attachment ambiguity.

of *properties* of nodes. The central idea is that representation types can be defined by picking out, from a phrase structure tree, subsets of nodes which share a certain property, which we will call the *Generator Property* (borrowing some terminology from Barker and Pullum (1990)).

We will illustrate this with reference to the tree for (5.1), given in Figure 5.4. The most fine-grained representation is the phrase-structure tree (marked PS tree in the diagram). It is the most fine-grained representation because it is composed of all the nodes. However, we can also pick out less fully-specified representations by picking out only *subsets* of nodes, which share a particular generator property. In the diagram, we have drawn a circle around all the nodes which we call *thematic assigner* nodes, and a square around all the nodes which we call *thematic receiver* nodes. Let us say that the set of *thematic* nodes consists of the union of the thematic assigner and thematic receiver nodes. Intuitively, the set of *thematic nodes* corresponds to the set of nodes which participate in confirmed thematic relations at some particular state. The relevant definitions, which we will go on to use in subsequent sections of this chapter, are as follows: (Note that the notion of "confirmation" referred to in these definitions will be discussed further in section 5.6, and can be ignored for present purposes).[5]

[5]i.e. "confirmed to have received/assigned" can be interpreted as "has received/assigned".

Thematic Nodes:

N is a *thematic node* at state S iff N is a *thematic receiver* at S or N is a *thematic assigner* at S.[6]

Thematic Receivers: N is a *thematic receiver* at state S iff N has been confirmed to have received a thematic role at or before S.

Thematic Assigners: N is a *thematic assigner* at state S iff it is the extended maximal projection of some head which has assigned at least one confirmed thematic role at or before S.

We can envisage various different levels of specification, according to which generator property we consider—for example, the phrase-structure tree consisting of all the nodes, the thematic tree consisting of all and only the thematic nodes, and the thematic assignment tree consisting of all and only the thematic assigner nodes. Notice that the trees become "flatter" as the membership of the relevant set of nodes becomes smaller.[7] It is possible to define monotonicity in terms of any one of these levels, or, in general, in terms of any arbitrary generator property.

What, then is the role played by the structural relations of dominance and precedence, and how does this role relate to the notion of generator properties?

The statement that A dominates B can be interpreted as a statement that domain A contains element B. On the other hand, a precedence relation between C and D can be interpreted as a statement that C and D are independent domains (i.e. neither one contains the other), since, by the theory of trees, a precedence relation between two nodes implies that the two nodes are not related by dominance[8]. The requirement that structural relations be updated monotonically can therefore be seen as a dynamic constraint on the construction of domains; that is, the constraint that both the membership (dominance) and the independence (precedence) of domains should be preserved.

When we move on to consider monotonic models defined in terms of some arbitrary generator property P, we simply use P to define the relevant notion of domain; that is,

[6]Recall that we are assuming that the "granularity" of states is word-by-word

[7]Of course, given a phrase structure tree labelled with the necessary thematic feature information, the "thematic trees" are purely derivative constructs, and do not need to be built separately by a parser. We depict them in the diagram (including terminal symbols for clarity) purely for expository convenience.

[8]The exclusivity condition of the theory of trees forbids two nodes from being simultaneously in a dominance and a precedence relation. See chapter 3, page 48 for a discussion of the theory of trees.

the domains whose membership and independence we wish to preserve consist of subtrees rooted in nodes of which P holds.

Thus, in a system defined by the generator property P, the *membership* of a domain is the set of nodes dominated by some P-node, while the *independence* of two domains is captured by a precedence relation between two P-nodes. So, for example, if we were to take the generator property to be that of *thematic nodes* (i.e. the union of thematic assigner and thematic receiver nodes), then, in figure 5.4, NP1 and NP2 would be independent domains (i.e. NP1 and NP2 are each thematic nodes, and a precedence relation holds between them), while, for example, NP3 would be a member of the domain of NP2, because NP2 dominates NP3.

We now move on to a general definition in which any arbitrary generator property can be used to define a domain. In the following definitions, the P-Dominator monotonicity constraint ensures that although each node can become a member of a new domain (where domains are defined in terms of generator property P), no node can cease to be a member of an existing domain. On the other hand, the P-Precedence partner monotonicity constraint ensures that once a pair of independent domains has been established, (i.e. two nodes, each with property P are in a precedence relation), that independence must be preserved.

P-Dominators:

The *P-dominators* for Y at state S (abbreviated as $D(Y,P)^S$) is the set of all nodes X, such that property P holds of X at state S, and X properly dominates Y at state S (where X *properly dominates* Y iff X dominates Y and $X \neq Y$).[9]

P-Dominator Monotonicity:

For each transition $S \Longrightarrow T$, and for each node X, $D(X,P)^S \subseteq D(X,P)^T$.

P-Precedence Partners:

The set of *Precedence Partners* at state S (written $PRP(P)^S$) is the set of all ordered pairs of nodes X and Y, such that property P holds of both X and Y at state S, and X precedes Y at state S.

[9]The use of the predicate "Properly Dominates" rather than "Dominates" allows a close convergence with the generalized command theory of Barker and Pullum (1990).

P-Precedence Partner Monotonicity:

For each transition $S \Longrightarrow T$, $PRP(P)^S \subseteq PRP(P)^T$

This generalized definition allows one to recreate Gorrell's constraints by simply setting the generator property P to the trivial property of being a node. This will have the effect of preserving dominance and precedence relations between all nodes. However, it also allows any number of alternative definitions to be made. In particular, if we can set the generator property to be the property of being a *thematic assigner node*, then the reanalysis which appears to be preferred in Italian and Japanese will not break the monotonicity constraint. Recall figure 5.4. Imagine that, as in de Vincenzi and Job's Italian experiments (1995), we have evidence for an initial on-line commitment to the low attachment of the relative clause (the reading in which the relative clause modifies *colonel*, as illustrated in the figure), followed by a final preference for the high attachment site (where the relative clause modifies *daughter*), and that this reanalysis is therefore of the "easy" kind. Let us say that this reanalysis occurs in the transition $S \Longrightarrow T$. Then, in Gorrell's system, the relative clause will lose at least one dominator (e.g. NP3) in the transition to state T, and the reanalysis will be incorrectly predicted to be "difficult". However, if we alter the generator property so that it picks out only the set of thematic assigner nodes, then the reanalysis will be correctly predicted to be unproblematic, since both dominators for the relative clause (i.e. both S and NP2) remain dominators for RC in both attachment positions in both S and T. It can also be seen that in cases where the preposition heading the PP assigns a thematic role, and the relative clause is attached low, this PP will be a dominator for the relative clause, preventing a subsequent monotonic reanalysis of the relative clause to the high attachment site.

5.4 Thematic Reception and Processing Difficulty

In the previous section, we claimed that, by taking thematic assignerhood as the relevant generator property, a monotonic model can account for the ease of reanalysis of modifier attachment that we were discussing. However, in this section, we will see that this results in a system which is too unrestrictive.

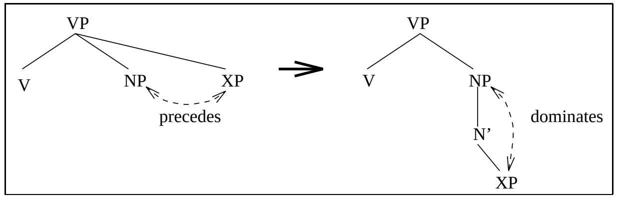


Figure 5.5:

5.4.1 Double Complement Verbs

The following sentences, involving verbs which can take two complements, have been referred to as "conscious" garden paths in the null context.[10]

(5.2) # The patient persuaded the woman that he was having trouble with to leave the room. (Gibson (1991), Pritchett (1992))

(5.3) # Mary saw the man with the binoculars through the telescope. (adapted from Gorrell (1995b))

Let us assume that these judgements are correct[11][12], and therefore that we want (5.3) and (5.2) to be unprocessable within the constraint of monotonicity. Assuming an initial argument attachment, of the *with* phrase in the case of (5.3), and of the *that* clause in the case of (5.2), then in both (5.3) and (5.2), the reanalysis involves the configurational change depicted in figure 5.5:

As we saw in chapter 3, Gorrell extended the standard D-theory formalism by applying monotonicity to *precedence* as well as *dominance* relations between nodes. This is in contrast to Weinberg (1993), who assumes that only precedence relations between terminal symbols are

[10] We will use the "sharp" symbol (#) to mark sentences which are typically thought to involve concsious processing difficulty, when it is convenient to do so.

[11] As in previous chapters, for the sake of argument, we simply assume that the intuitive judgements offered by previous researchers are correct. We do not deny the need to back up these judgements empirically.

[12] Clifton (1995) describes an eye tracking experiment which investigated garden path sentences of the type illustrated in (5.2). Though the experiment was intended to investigate a separate issue, Chuck Clifton informs us (personal communication) that, in addition to garden path effects found in the eye-movement data, response accuracy to comprehension questions was extremely low for the dispreferred relative clause continuation, in comparison with the preferred complement clause continuation (34 vs. 85 % correct). This indicates that reanalysis is extremely hard, and often impossible in this type of garden path. Note that the sentences in this experiment were presented in the null context.

subject to the monotonicity criterion. In figure 5.5, there is no constituent that dominates the XP before reanalysis which fails to dominate it after reanalysis—the revision may be carried out simply by adding the relations `dom(NP,XP)` and `dom(N′,XP)`. However, as Gorrell says, the difficulty of (5.2) can be captured if we make the assumption that the preservation of precedence relations (as well as dominance relations) must also be added to the monotonicity criterion. Once this is done, the difficulty of (5.2) is correctly accounted for; NP *precedes* XP before reanalysis, but NP dominates XP *after* reanalysis, and, according to the theory of trees, no two nodes can be in both a dominance and a precedence relation. In terms of the model given in chapter 4, this is captured by the fact that, in both (5.2) and (5.3), the NP ceases to be accessible as soon as the XP is attached as a daughter of VP.

Now consider the consequences of assuming a monotonic system in which the relevant generator property is that of being a thematic assigner node, as we suggested in the previous sections. In neither case does the NP direct object have the relevant generator property, because in neither (5.2) nor (5.3) does its head noun assign a thematic role. This means that NP and XP will not be precedence partners, and thus the reanalysis depicted in figure 5.5 will not violate monotonicity, an undesirable result. It seems that what is relevant in these cases is related to the *reception* rather than the *assignment* of thematic roles. Specifically, in both (5.2) and (5.3), NP and XP are co-recipients of thematic roles from the verb before reanalysis (i.e. independent *thematic receiver nodes*, but this is not so after reanalysis.[13]

5.4.2 Pritchett (1988): Thematic Reception

In fact, a constraint on reanalysis defined in terms of *thematic reception* has already been proposed; it is the *Theta Reanalysis Constraint* of Pritchett (1988), which we discussed in chapter 2, and we repeat below:

> **Theta Reanalysis Constraint:** Syntactic reanalysis which interprets a θ-marked constituent as outside of its current θ-domain is costly.
>
> **θ-domain:** α is in the γ θ-domain of β iff α receives the γ θ-role from β or α is dominated by a constituent that receives the γ θ role from β.

[13]It might be suggested that the difficulty of (5.2) and (5.3) is related to the fact that the constituent XP is a thematic assigner node in both cases (a PP and a clause respectively). However, a similar "conscious garden path" effect has been claimed for the following sentence in which the equivalent constituent, *the report* is not a thematic assigner node: *They gave the man the report criticised a demotion.*

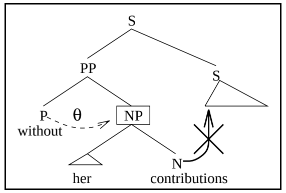


Figure 5.6: Informal illustration of Pritchett's Theta Reanalysis Constraint. The "box" surrounding the NP represents the fact that this NP has received a thematic role.

As an example of the Theta Reanalysis Constraint in action, consider (5.4), an example of a "conscious garden path" from Pritchett's paper:

(5.4) # Without her contributions failed to come in.

Pritchett assumes that the NP *her contributions* receives a thematic role from the preposition *without*. Because *contributions* is dominated by this NP, any reanalysis which "reinterprets" *contributions* into a position from which it is no longer dominated by this NP, is predicted to be costly. We illustrate this in Figure 5.6.

It can be seen that, if we define a monotonicity constraint with *thematic reception* as the generator property, then the difficulty of (5.2) and (5.3) can be accounted for. This is because, referring again to figure 5.5, NP and XP will both have the generator property, because they will both receive thematic roles from the verb. This will mean that the two constituents will be precedence partners before reanalysis takes place. The reanalysis depicted in figure 5.5 will result in the breaking of this precedence partner relation, and will therefore be ruled out by the monotonicity constraint.

5.4.3 Thematic Reception and Thematic Assignment

On the basis of the data discussed in section 5.4.1, we have argued that the relevant notion of monotonicity cannot be motivated solely in terms of *thematic assignment* nodes. However, it is also true that, on the basis of the modifier attachment data discussed in section 5.2, monotonicity cannot be defined solely in terms of *thematic receiver* nodes. To see this, consider again figure 5.4 on page 110. If monotonicity were defined with thematic *receiverhood* as the generator property, then the reattachment of the relative clause from the low site to the high site would violate monotonicity, since NP3, being a thematic receiver, would be a dominator for the relative clause before reanalysis, but would fail to be after reanalysis.

Let us step back and consider the position in which we find ourselves. On the basis of data concerning the reanalysis of modifier attachment, we have argued that monotonicity should be defined in terms of *thematic assignment*. However, on the basis of data concerning the reanalysis of *argument* attachment, we have argued that monotonicity should be defined in terms of *thematic reception*. Hence, it appears to be necessary to employ a different notion of monotonicity depending on the *type* of constituent which is being reanalysed; that is to say, the thematic dominators for some constituent C will have to depend partly on the features of C itself; that is to say, the notion of monotonicity required is *relational* rather than *functional*. In the next section, we will introduce a method of defining monotonicity in terms of *relations*, which will allow us to capture the required two-level constraint.

5.4.4 Generalizing Monotonicity in Terms of Relations

The second, more powerful way in which we can generalize monotonicity is in terms of *relations*. In section 5.3.2, where monotonicity was defined in terms of *properties*, the P-dominators for a node N were defined exclusively in terms of the generator property P, along with the predicate *properly dominates*. By contrast, when we define monotonicity in terms of *relations*, the membership of the set of dominators of N may depend partly on features of N itself. This allows us to make different types of nodes sensitive to different types of domains. As in section 5.3.2, this method of generalizing monotonicity constraints is inspired by the Generalized Command theory of Barker and Pullum (1990).

We now move on to the definitions. In what follows, we use the the abbreviation $R(X,Y)^S$ to stand for "The relation R holds between X and Y at state S". We will refer to R as the *Generator Relation*.

R-Dominators:

The *R-dominators* for Y at state S (abbreviated as $D(Y,R)^S$) is the set of all nodes X such that $R(X,Y)^S$, and X properly dominates Y at state S.

R-Dominator Monotonicity:

For each transition $S \Longrightarrow T$, and for each node X, $D(X,R)^S \subseteq D(X,R)^T$.

R-Precedence Partners:

The set of *R-Precedence Partners* at state S (written $PRP(R)^S$) is the set of all ordered pairs of nodes X and Y, such that $R(X,Y)^S$ and $R(Y,X)^S$, and X precedes Y at state S.[14]

R-Precedence Partner Monotonicity:

For each transition $S \Longrightarrow T$, $PRP(R)^S \subseteq PRP(R)^T$

In section 5.3.2, we defined a particular instantiation of generalized monotonicity by specifying a value for the generator property variable P. Similarly, when we define a model in terms of a relational monotonicity constraint, we have to specify the value of the generator relation R. In the next section, we will define *visibility* as a generator relation.

5.5 Thematic Monotonicity

In what follows we will introduce the *thematic monotonicity* constraint, which will be defined in terms of the *visibility* relation. Intuitively, the *visibility* relation will pick out, for each node, which other nodes in the tree are relevant to the constraints on its reanalysis.

We begin by introducing some terminology.

5.5.1 Thematic Nodes

We have already introduced the notion of *thematic nodes*. Here we repeat the definition:

Thematic Nodes:

N is a *thematic node* at state S iff N is a *thematic receiver* at S or N is a *thematic assigner* at S.

Thematic Receivers: N is a *thematic receiver* at state S iff N has been confirmed to have received a thematic role at or before S.

[14] Note that, for the sake of generality, we have defined R-Precedence Partners "symmetrically", assuming no difference in priority between the left and right members of a pair of nodes in a precedence relation. An anti-symmetric definition would also be possible, where only one of $R(X,Y)$ or $R(Y,X)$ would have to hold. This would require independent justification of the directionality of the relation.

Thematic Assigners: N is a *thematic assigner* at state S iff it is the extended maximal projection of some head which has assigned at least one confirmed thematic role at or before S.

5.5.2 Visibility

The thematic monotonicity constraint will be defined *relationally*, so that the set of thematic nodes relevant to some constituent C will depend on some of the properties of C itself. We will begin, then, by defining the relation of *visibility*[15] which will define this set of "relevant" thematic nodes for any given constituent C. Recall that Pritchett's TRC and Frazier and Clifton's *Construal Principle* differ, not only with respect to the reception versus assignment of thematic roles, but also with respect to the features of the constituent to which the respective principles apply; specifically, Frazier and Clifton's principle applies to "Secondary Phrases", while Pritchett's principle applies to "Theta marked" constituents. We will preserve the substantive content of this distinction in the visibility relation which we define below, by making modifiers sensitive only to *thematic assigner* nodes, while non-modifiers are sensitive to *all* thematic nodes (i.e. both thematic receivers and thematic assigners). This means that, in the general case, modifiers will have more freedom to be revised than other nodes.[16]

Consider figure 5.4 on page 110. Because the relative clause, RC, is a modifier (i.e., in Construal theory, it would be considered as a "Secondary Phrase"), we are going to say that only *thematic assigner* nodes are visible to RC. This will mean that the RC can be "raised" past a thematic receiver node, such as NP3, without violating monotonicity, because such nodes are not dominators for RC. However, the relative clause may *contain* nodes which are non-modifiers, so the visibility relation must allow all nodes dominated by RC to inherit RC's "blindness" to NP3. Hence, the visibility relation will define modifiers to act as what may intuitively be thought of as a "visibility filters" to nodes inside them—that is, in the case of RC in figure 5.4, the only nodes *outside* RC that are visible to nodes *inside* RC are thematic assigner nodes. So, for example, because NP3 is not a thematic assigner node, it is not visible

[15] We are aware of a possible confusion caused by using the term *visibility*, which has been used in many other proposals, both in theoretical syntax and psycholinguistics, each with reference to a slightly different concept. *Visibility* seemed the most appropriate name for the relation we define, and we hope the reader will not be confused by our use of it here.

[16] We believe that it is reasonable to treat modifiers separately in this way, since modifiers, by definition, never receive thematic roles, and can therefore be hypothesised to be insensitive to thematic receiver nodes in general.

to RC, or to any nodes within RC. On the other hand, both S and NP2 are thematic assigner nodes, and so *are* visible to RC and to nodes within RC. We will say that *all* thematic nodes, that is both assigners and receivers, are visible to non-modifiers, unless they are "filtered out" by a modifier in this way.

The "filtering" behaviour of modifiers gives the visibility relation a rather complicated and stipulative appearance. This is probably a consequence of the fact that we are defining the thematic monotonicity constraint in terms of the generalized "recipe" provided in section 5.4.4, and an equivalent, more natural definitions are possible. When we consider the computational implementation described in chapter 6, this "filtering" can be seen to follow naturally from the strictly local manner in which information concerning the global configuration is inherited from mother to daughter through the tree.[17] It may also be noted that when certain command constraints are defined in terms of relations in the the theory of Barker and Pullum (1990), the resulting relations can often be complex. However, what is gained by defining constraints in this way, whether monotonicity constraints or command relations, is generality, which allows us to analyse the proposed constraints according to a common framework.

The definition of "visibility", then, is as follows:

Visibility

1. A is *visible* to a modifier M iff A is a thematic assigner node.[18]
2. A is *visible* to a non-modifier N iff

   either there is no modifier that dominates N but not A, and A is a thematic node (i.e. a receiver or assigner).

   or there is a modifier that dominates N but not A, and A is a thematic assigner node.

Now, instead of insisting that dominance and precedence relations should be preserved between all nodes, as is the case in other D-theory based models (Sturt and Crocker (1996),

[17] Readers familiar with 1980's Chomskyan syntax will also recognize some parallels between the "filtering" behaviour of modifiers proposed here, and the idea, put forward in Barriers (Chomsky (1986a)), that, under certain conditions, non-theta-marked constituents serve to block the formation of various linguistic dependencies from external nodes to descendants of the category in question.

[18] This and the remainder of the definitions in this section are intended to be relativized to *states*. Strictly speaking, this particular clause of the definition of visibility should read "A is *visible* to a modifier M at state S iff A is a thematic assigner node at state S". We omit mentioning states to avoid cluttering the definitions.

Gorrell (1995b)), the thematic monotonicity constraint is is going to insist only on the preservation of dominance and precedence between nodes related by *visibility*.

5.5.3 Thematic Dominators and Thematic Precedence Nodes

What we will call *Thematic Dominators* and *Thematic Precedence Partners* are simply the result of substituting the *visibility* relation in the appropriate definition from section 5.4.4.

Thematic Dominators:

> A is a *thematic dominator* for B iff A is visible to B and A properly dominates B.[19]

Recalling figure 5.4, the thematic dominators for the relative clause consist of NP2 and S, while the thematic dominators of the N$'$ immediately dominating the relative clause consist of NP3, NP2 and S.

Thematic Precedence Partners:

> A is a *thematic precedence partner* for B iff A and B are visible to each other and A and B are in the precedence relation.[20]

Again, recalling figure 5.4, NP1 is a thematic precedence partner for NP3, because the two nodes are visible to each other, and NP1 precedes NP3. However, NP1 is not a thematic precedence partner for RC, because the two nodes are not visible to each other (specifically, although RC is visible to NP1, NP1 is not visible to RC).

5.5.4 The Thematic Monotonicity Constraint

The thematic monotonicity constraint, then, is defined as follows:

The Thematic Monotonicity Constraint

1. A node can gain new thematic dominators, but cannot lose existing ones (that is, for each node, the set of thematic dominators is monotonic increasing).
2. A node can gain new thematic precedence partners, but cannot lose existing ones (that is, the set of pairs of precedence partners is monotonic increasing).

[19] A properly dominates B iff A dominates B and $A \neq B$.
[20] Here "in the precedence relation" means A precedes B or B precedes A.

5.6 Delayed Commitment

In this section, we will discuss the notion of *confirmation* which is mentioned in the definition of thematic assigner and receiver nodes.

Recall the two "difficult" garden path sentences discussed above:

(5.2) # The patient persuaded the woman that he was having trouble with to leave the room. (Gibson (1991), Pritchett (1992))

(5.3) # Mary saw the man with the binoculars through the telescope. (adapted from Gorrell (1995b))

Despite the structural similarity, it has also been claimed that the following sentences do *not* cause the conscious garden path effect:

(5.5) Mary saw the man with the moustache.

(5.6) The doctor told the woman that was having trouble with him to leave the room. (Gibson (1991), Pritchett (1992))

Note that, if we assume a processor which makes attachments on an incremental word-by-word basis, as well as an initial preference for the VP attachment in both cases, then, in constituent structure terms, both (5.3) and (5.2) on the one hand, and (5.5) and (5.6) on the other, involve an identical configurational change, namely that depicted in figure 5.5 (see page 114). They only differ in the point at which disambiguation occurs.

Let us briefly summarize the dilemma in which we find ourselves. Assume first that we want our model to be incremental, in the sense that each word is incorporated into a connected structure as soon as it is encountered. Let us also assume that a node which is eligible to be a thematic node (i.e. either an assigner or a receiver) will be confirmed as such as soon as it is created. In this case, (referring again to figure 5.5, page 114), in both (5.5) and (5.6), the second complement of the verb (labelled "XP" in figure 5.5) will be confirmed as a thematic receiver as soon as the first word of XP is read. Therefore, NP and XP will be thematic precedence partners, and the reanalysis will be (falsely) predicted to be difficult.

Now assume that we do not care if we relax the assumption of incrementality. In this case, we might propose a processor which, for some reason, delays the decision to attach XP, in such a way that this constituent remains unattached at the point of disambiguation in the "easy" cases (i.e. (5.3) and (5.2)), but is already irrevocably attached by the time of disambiguation in the "difficult" cases (i.e. (5.5) and (5.6)).

However, this solution is not satisfactory either—it is not consistent with the abundant evidence for incremental, word-by-word processing, which we discussed in chapter 3. In particular, it predicts, not only that (5.5) and (5.6) should not involve conscious reanalysis, but it makes the much stronger prediction that they should involve no reanalysis at all. However, contrary to this, there *is* evidence for reanalysis in both of these constructions (see Rayner, Carlson, and Frazier (1983) for (5.5) and Mitchell, Corley, and Garnham (1992) for (5.6)). This means that our model must not only represent the initial commitment to the argument analysis in these sentences, but also, if we accept the intuitive distinction at issue, be capable of revising in the case of (5.5) and (5.6) but not in the case of (5.3) and (5.2). More specifically, it appears that we need to to be able to encode more than one level of commitment into the model—an immediate, but relatively weak commitment, which can be relatively easily broken at the point of disambiguation in (5.5) and (5.6), as well as a firmer, but more delayed commitment, which causes the processing trouble in (5.3) and (5.2). Below, we will show that this notion of delay can be expressed quite naturally in the monotonic system proposed in this chapter.

As we have mentioned above, there is a great deal of evidence that attachment is basically a word-by-word incremental process, and does not require the presence of a grammatical head. However, it is possible that, while the level of commitment corresponding to syntactic *attachment* is not head-driven, some higher level of commitment *is* head driven. This assumption has been employed in some recent models. For example, Ferreira and Henderson (1991) suggest that the assignment of a thematic role can only be achieved once the head of the constituent receiving the role has been read.

Though Ferreira and Henderson deal with phenomena which are rather different from those we describe here, we will incorporate a similar proposal into the present model. In our case, we do not argue that the presence of a head is necessary for thematic role assignment or interpretation. Rather, we see the presence of a head as a prerequisite for the *confirmation* of a node as thematic. That is to say, the presence of a head in the input may trigger the creation of one or more thematic nodes, and the effect of this will be that certain choice points

representing alternative parses will be made inaccessible to monotonic reanalysis processes after confirmation has taken place.

The idea of making commitments at various points during the processing of a string has precedents in a number of previous proposals, dating back at least as far as the *Sausage Machine* (Frazier and Fodor (1978)), and is also seen clearly in Gibson (1991). Bounding reanalysis in this way has clear computational advantages in terms of preserving working memory, but in the context of the current proposal, it is also directly linked to the preservation of certain aspects of thematic structure. We would expect an adaptive processor to make such commitments at maximally informative points in the input, and heads of phrases would seem to be a natural choice.

Recall that the definition of thematic nodes includes the stipulation that they be *confirmed*. Below we repeat the definition:

Thematic Nodes:
N is a *thematic node* at state S iff N is a *thematic receiver* at S or N is a *thematic assigner* at S.

> **Thematic Receivers:** N is a *thematic receiver* at state S iff N has been confirmed to have received a thematic role at or before S.
>
> **Thematic Assigners:** N is a *thematic assigner* at state S iff it is the extended maximal projection of some head which has assigned at least one confirmed thematic role at or before S.

We define thematic confirmation as follows:

Thematic Confirmation:
Two nodes A and R are *confirmed* as Thematic Assigner and Thematic Receiver respectively iff

i. R is in a position from which it may receive a thematic role from the head of A.
ii. The head words of both A and R are instantiated.
iii. The resulting thematic dependency is semantically acceptable.

Consider the simple phrase structure tree for the sentence *Mary saw the man with the binoculars*, given in figure 5.7:

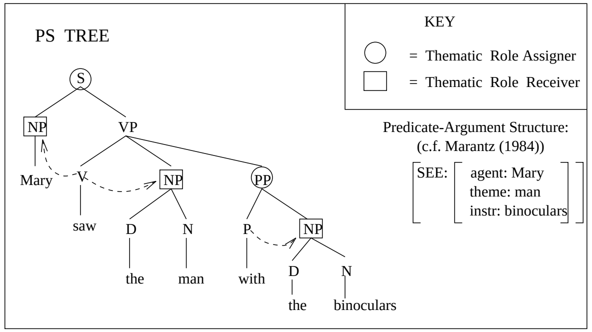


Figure 5.7:

We are assuming a notion of thematic structure which is similar in spirit to that of Marantz (1984). That is, we assume that each argument of a verb is licensed to appear in that verb's predicate-argument structure by receiving a thematic role.[21] This thematic role may be assigned either by the verb itself, in which case the argument is referred to as a *direct* argument, or otherwise by a theta assigning head which is distinct from the verb, such as an adposition[22], in which case the argument is called an *indirect* argument of the verb.

In the particular case of the sentence illustrated in figure 5.7, we assume that *the binoculars* occupies the INSTRUMENTAL slot in the predicate-argument structure of SEE, but in order to occupy this position, it has to be assigned its instrumental thematic role by *with* (that is, *the binoculars* is an *Indirect Argument* of *saw*). This makes the PP a *Thematic Assigner* node, as it is the maximal projection of *with*, the word that assigns the instrumental role. The other thematic assigner node is the S, which is the extended maximal projection of *saw*, the word which assigns the AGENT role to *Mary* and the PATIENT role to *the man*. Finally, the thematic receivers are the three NPs, two of which receive their thematic roles from the verb, and one from the preposition.

Let us reconsider our example. Figure 5.8 represents the state at which the processor has successfully parsed the substring *Mary saw the man with the*.

[21]Marantz calls these "semantic roles".

[22]By "adposition" we mean either a preposition or a postposition.

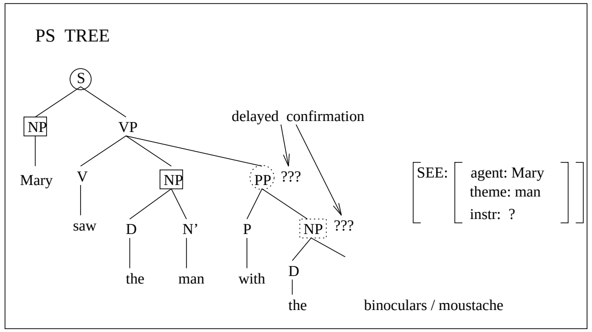


Figure 5.8:

The potential recipient of the instrumental role (i.e. the NP enclosed in the dotted box) has not yet been confirmed as a thematic node, since its head is not yet instantiated. Similarly, the PP which is a potential thematic assigner node has not yet been confirmed, as the head of its Receiver Node has not yet been read. If the input continues with the head *binoculars*, then the two nodes may be confirmed. This confirmation results in a precedence relation between two thematic nodes; the PP and the preceding NP dominating *the man*. This means that the PP cannot subsequently be reanalysed into a position dominated by this NP without violating thematic monotonicity (the precedence relation would have to be broken). If the input continues with *moustache*, however, we assume that the implausibility of a moustache as an instrument prevents the confirmation of the instrumental role from taking place, and the PP may therefore be reanalysed as a modifier of the NP.

Recalling the two examples (5.3) and (5.5) repeated below, we now have an explanation for the intuitive difference.

(5.3) # Mary saw the man with the binoculars through the telescope.

(5.5) Mary saw the man with the moustache.

In (5.3), once the two thematic nodes in the PP *with the binoculars* have been confirmed, these nodes become *visible* to, and therefore thematic precedence partners of, the preceding NP headed by *man*. The PP cannot now be reanalysed into a position dominated by this NP, as that would result in a thematic precedence partner relation becoming a thematic dominator relation, against the theory of trees.

5.7 Coverage and Predictions

Thus far, we have motivated the model based on a relatively small number of constructions. We now give an idea of the wider coverage of the model.

5.7.1 Comparison with Pritchett (1988)

In general, given the importance attached to receivers of thematic roles, the model will share many predictions with Pritchett (1988). We believe that the model has empirical advantages over Pritchett's in two respects. First, because Pritchett's model was based on the grammatical notion of *licensing*, unlike the present proposal, it made no concrete predictions concerning the attachment and reanalysis of modifiers; constituents which have no clear relation with licensing principles. Second, initial attachment in Pritchett's model was entirely head-driven, which is problematic with respect to the abundance of evidence for incremental, word-by-word processing. By contrast, we assume that initial attachment is word-by-word incremental, but that the importance of heads is reflected in a head-driven *commitment* to thematic structure. Note, however, that in its present form, the model will not extend to cover the phenomena which motivated Pritchett to reformulate his reanalysis principle as the On Line Locality Constraint (OLLC), in Pritchett (1992). In fact, we believe that the OLLC does not hold the empirical advantages that Pritchett assumed. Specifically, the OLLC was designed to account for the fact that, intuitively, both (5.7) and (5.8) are equally easy:

(5.7) John gave her earrings for Christmas.

(5.8) John gave her earrings away to charity.

The "theta attachment" principle of Pritchett's model predicts that *her* and *earrings* are initially attached as the indirect object and direct object of *gave* respectively (we will call this the *dative reading*. However, on the assumption that the dative reading is initially preferred, (5.8) requires *her earrings* to be restructured as a single NP, the direct object of *gave*, (we will call this the *possessive reading*, because *her* is interpreted as a possessive. This violates the

TRC, because the thematic role originally assigned to *earrings* will have to be removed. On the other hand, Pritchett's revised principle, the OLLC predicts the reanalysis outlined above to be easy, because it allows the source position of a node to *govern* its target position. However, the OLLC only solves the problem on the assumption that the initial preference for the dative reading is correct. If *earrings* is initially in accordance with the *possessive* reading, then both the TLC and the OLLC will predict major processing difficulty associated with (5.7). In fact, there is no evidence for the initial dative preference that Pritchett assumed, and, though inconclusive, some data which does exist for this construction in English shows evidence for an initial preference for the *possessive* reading.[23] This preference for the *possessive* reading has also been found in a completion study on the equivalent construction in German (Bader, 1996), but Bader reports that in a self-paced reading study, a strong preference for the possessive reading is dependent on the presence of a focus particle; a result which Bader explains in terms of the difficulty of revising intonational structure. In the thematic monotonicity model which we have introduced here, we could only explain the apparent ease of processing in both (5.7) and (5.8) if we allowed an extra degree of underspecification, perhaps due to the lexical ambiguity of *her*. However, this would require independent justification, and, if we were to provide an account of Bader's German data, would also require the introduction of focus structure into the model. We leave this as a problem for future research.

5.7.2 Comparison with Frazier and Clifton (1996)

As far as modifier attachment is concerned, given the importance attached to *assigners* of thematic roles, the model will share many predictions with Construal Frazier and Clifton (1996). One difference, however, is that while Construal Theory "associates" modifiers into a thematic domain without committing at the syntactic level to any one particular site within the domain, our model makes an initial determinate attachment at the *constituent structure* level, while the *thematic* domain demarcates the locus of possible alternatives to this attachment choice. As we have seen in the discussion in section 5.2, data currently available do not allow us to discriminate between these two choices.

However, there is one respect with which Thematic Monotonicity, as formulated in this chapter, is superior to Construal, and that concerns left branching attachment of modifiers.

[23] Chuck Clifton, personal communication.

Consider the following Japanese NP:

(5.9) [$_{RC}$ barukonii ni iru] joyuu no mesitukai.
balcony LOC exists actress GEN servant
"The servant of the actress who is on the balcony"

Recall that the Construal Principle states that non-primary phrases are "Associated" into the current thematic processing domain, which is defined as "the extended maximal projection of the last theta assigner". This definition makes sense for right branching structures, but it does not make sense for head-final left branching structures of the type illustrated in (5.9). At the point where *joyuu* is reached, it becomes possible that the clause labelled RC is a relative clause modifying *joyuu*. However, at this point, the last thematic assigner is the main verb of the relative clause, which means that the Current Thematic Processing Domain is the relative clause itself. Clearly, it is implausible to assume that a relative clause should be "associated" to itself. In contrast, consider the Thematic Monotonicity Constraint. This requires that the relative clause should be able to gain new thematic dominators, but not lose existing ones; that is, that the set of thematic assigner nodes properly dominating RC should be monotonic increasing. We assume that RC is initially attached as a modifier of *joyuu*, as suggested by Kamide and Mitchell's (1996) on-line study. At this point, the set of thematic dominators for RC is the empty set. Subsequent input will result in additions to this set of thematic dominators for RC. Specifically, the NP headed by *mesitukai* will become a member of this set, because *mesitukai* assigns a thematic role to the NP headed by *joyuu*, and if this NP itself becomes incorporated into a larger left branching structure through further input, then more dominators may well be added. This will allow the relative clause to "climb" to a position where it modifies *mesitukai*, without losing any thematic dominators.

5.7.3 Three Site Relative Clause Ambiguities

One phenomenon which we have not discussed so far is the pattern of preferences in relative clause ambiguities with three possible attachment sites, such as the following NP from the study reported by Gibson et al. (1996):

(5.10) the design on the kite above the house that was small but very beautiful.

The relative clause in (5.10) may be attached at any one of three sites; it may modify the *low* noun (*house*), the *middle* noun (*kite*), or the *high* noun (*design*). Using a cumulative grammaticality judgement task, Gibson et al. (1996) showed a preference ordering of *Low* > *High* > *Mid*, for disambiguated versions of (5.10) in both Spanish and English. This was explained in terms of two factors; *recency*, which favours attachment to recently processed material, and *predicate proximity*, which favours attachment as close as possible to the head of a predicate phrase, giving an advantage to the *high* site in examples such as (5.10).[24]

In terms of the thematic monotonicity model presented in this chapter the *high* and *mid* attachments must be derived via reanalysis, assuming an initial low attachment.[25] This means that the preference for the high site over the middle site must be seen as an effect of reanalysis. Let us consider some questions about how reanalysis might proceed in such examples. Gibson et al's materials have the schematic form indicated in (5.11):

(5.11) NP_1 P_1 NP_2 P_2 NP_3 RC

An examination of Gibson et al's materials reveals that, P_2 is instantiated by a thematic preposition in all but three of the eighteen materials. This means that, assuming the high and low attachment conditions are derived via reanalysis, the reanalysis required in the majority of Gibson et al's materials does not preserve thematic monotonicity, since the relative clause instantiating RC must be moved beyond the maximal projection of P_2. This suggests that predicate proximity, or some similar principle, must be seen as guiding non-monotonic reanalysis processes. However, given thematic monotonicity, we might assume that the preference for the high site over the middle site might be reversed in cases where P_2 is *non*-thematic. In fact, we have some preliminary evidence for this, which we describe below.

Pilot Questionnaire Study

In this questionnaire study,[26] we examined twelve sentences similar to (5.12), where the NP headed by *room* instantiates the schema given above in (5.11).

[24] Note that in the case of bare NPs, predicate proximity applies to a *predicted* predicate phrase. In the case of (5.10), the predicted predicate phrase takes (5.10) as its subject.
[25] Although a high attachment preference for Spanish is well attested in two-site examples (Cuetos and Mitchell, 1988), Gibson et al's results show a low preference for three-site examples.
[26] Because this is only a preliminary investigation, we do not give a full report of this study.

(5.12) The young couple liked the room with the painting of the horses that was really colourful.

In all materials, the relative clause was made incompatible with the most recent noun, N_3, using number marking, leaving a choice between the middle site N_2 and the high site N_1. In all materials, P_1 was instantiated by a thematic preposition and P_2 by the non-thematic preposition *of*, so that, assuming an initial low attachment of the relative clause, a final interpretation with RC modifying N_1 would involve non-monotonic reanalysis, while a final interpretation with RC modifying N_2 would involve monotonic reanalysis. Subjects were asked to choose between two possible interpretations corresponding to the middle attachment (*the painting was really colourful*) and the high attachment (*the room was really colourful*). There was a preference for the middle attachment over the high attachment (65% vs. 35%), which was significant by both subjects and items.

The pilot study was incomplete in many ways (for example, the plausibility of the two grammatical readings was not controlled, and we did not compare (5.12) with three-site examples in which P_2 is thematic). However, the results that we have suggest that thematic structure constrains the final interpretation of three-site relative clause ambiguities. It should be noted that the combined principles of predicate proximity and recency given in Gibson et al. (1996) cannot account for for this effect, since predicate proximity applies solely to *verbal* predicates.

5.7.4 "Easy" Revisions

Intuitively, the "unproblematic" reanalyses in this model correspond to revisions which do not cause drastic changes to the thematic tree. These may include cases in which the thematic tree does not change at all, as in the Italian and Japanese examples we discussed in section 5.3.1. Also included are cases in which a new thematic node is "inserted" into the domination path between two existing thematic nodes. An example of this is the well-known class of NP/clausal complement ambiguities such as (5.13) below, in which a new thematic assigner node representing the complement clause is inserted in the domination path between the root clausal node and the NP thematic receiver node covering *the ancient inscription*.

(5.13) The scholar understood the ancient inscription had been written by the Romans.

In general, all reanalyses which can be achieved through the *tree-lowering* operation given in chapter 4, will also be possible in the model given here. Also predicted unproblematic are cases in which a constituent is "lowered" into the domination domain of an immediately preceding thematically confirmed phrase, just as long as the lowered constituent, or any constituent which it dominates, has not been thematically confirmed at the time that the lowering takes place. Examples are (5.5) and (5.6) repeated below:

(5.5) Mary saw the man with the moustache.

(5.6) The doctor told the woman that was having trouble with him to leave the room.

We have dealt with (5.5) in the main body of the text. The example (5.6) is predicted to be unproblematic if we adopt Gibson (1991)'s assumption that *told* assigns the proposition role not to the (CP) projection of the complementizer but to the (IP) tensed clause headed by *was*. On this analysis, the embedded IP will still be unconfirmed as a thematic receiver before the disambiguating signal (i.e. the absence of the an overt subject in the embedded clause) is detected, and therefore reanalysis of the embedded clause as a relative is possible.[27]

5.7.5 "Difficult" Revisions

Revisions which are predicted to be problematic include cases in which a constituent loses one or more of its thematic dominators, as is the case in the well known example below:

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

In (5.14), assuming an initial low attachment preference for the NP *the sock*, this NP has to be raised past the subordinate clause node in order to gain the globally correct reading. As this clausal node is a thematic dominator for the NP (it is the extended maximal projection within which *mending* assigns its thematic roles), the reanalysis involves the NP "losing" this clausal node as a thematic dominator.

Also predicted to be difficult are cases in which a precedence relation between two confirmed thematic nodes has to be broken, as in examples (5.3) and (5.2) repeated below:

(5.3) # Mary saw the man with the binoculars through the telescope.

[27] Note that this analysis would also be in line with Head Driven Phrase Structure Grammar (Pollard and Sag (1994)), in which complementizers are treated as *markers* rather than heads.

(5.2) # The patient persuaded the woman that he was having trouble with to leave the room.

There is a class of revisions for which our model is probably over-constrained. That is cases like the following, where the relative clause, assuming an initial low attachment, has to be reanalysed to a position outside the domain of the thematic assigning preposition *with*.

(5.15) They laughed at the girl with the boyfriend who looked pleased with herself.

Though, as we mentioned in section 5.3.1, there is evidence that this type of ambiguity is influenced by whether or not the preposition assigns a thematic role, it is unclear whether examples such as (5.15) should be thought of as conscious garden paths, as we predict should be the case, on the assumption that violations of thematic monotonicity correspond precisely to the class of conscious garden paths.[28]

5.7.6 Grades of Difficulty

In this paper, we have divided the data into two categories according to the simple binary distinction between those reanalyses which violate thematic monotonicity and those which do not. However, we do not wish to claim that there are not grades of difficulty *within* either of these categories. Consider the following minimal pair from Fodor and Inoue (1994):

(5.16) a. She put the candy in her mouth on the table. HARDER

b. She put the candy in her mouth onto the table. EASIER

In terms of the model proposed here, we predict both of the examples in (5.16) to be hard, in the sense that they both involve a violation of thematic monotonicity, for essentially the same reason as in example (5.3), section 5.4. However, Fodor and Inoue point out an intuitive distinction between the more difficult (5.16.a) and the less troublesome (5.16.b). Fodor and Inoue motivate this in terms of a difference in informativity between a weaker, pragmatic error signal in (5.16.a) (i.e. the implausibility of a mouth being on a table), and a stronger, syntactic error signal in (5.16.b) (i.e. the impossibility of *into* as a modifier of *mouth*). However, on the fairly natural assumption that the difficulty of reanalysis correlates with the *number* of violations of thematic monotonicity required for recovery, the intuitive difference between

[28]For example, Gibson (1991) treats similar examples as not involving conscious difficulty.

(5.16.a) and (5.16.b) can be explained in the present framework. Assume that, for lexical reasons, the processor never attempts to attach *into* as a modifier of *mouth* in (5.16.b), but does attempt to attach *in* as a modifier of *mouth* in (5.16.a). Then, the reanalysis of (5.16.a) involves not only breaking the precedence relation between *the candy* and *in her mouth*, but also removing *on the table* from the thematic assignment domain of *in her mouth*, that is, two violations of thematic monotonicity. By contrast, (5.16.b) involves only the former violation. A similar account is possible for another of Fodor and Inoue's example pairs given below in (5.17), on the assumption that the processor does attempt to attach *the story* into the *that* clause in (5.17.a), but does not attempt to attach *not* into the *that* clause in (5.17.b):

(5.17) a. They told the boy that the girl met the story. HARDER

b. They told the boy that the girl met not to go home. EASIER

5.8 Conclusion

In this chapter, we have proposed two basic revisions to the theory of monotonic parsing which, as we hope we have demonstrated, allows a more flexible and psychologically natural model; first, we propose that monotonicity should be defined over *thematic* rather than *constituent* structure, allowing us to incorporate insights from other models making substantial use of thematic structure (Pritchett (1988), Frazier and Clifton (1996)). Second we have proposed that commitment is delayed, and *head-driven* at the thematic level, but immediate and incremental at the phrase-structural level. This allows us to capture a wider range of "mild" reanalyses which demonstrably occur in sentence comprehension, but which appear to cause readers little trouble.

In order to gain a clear idea of how the model proposed here fits into the monotonicity framework, it may be instructive to refer back to figure 3.1, page 44. Recall that in monotonic models, the *description* comprises a monotonic increasing set of structural relations. In the present model, this monotonic set of structural relations includes *thematic dominator* relations and *thematic precedence partner* relations. The "possible trees" consist of the infinite set of phrase structure trees compatible with these relations at each state. Finally, the "default tree" is "ready made" at each state—it is simply the phrase structure tree. This contrasts with the model described in chapter 4, where the default tree at any state is simply the "closed world" interpretation of the description at that state, and no further information is required in order to interpret the structure.

An interesting question which we have not addressed up to this point concerns the processing status of constituent structure-based monotonicity (cf. chapter 4) in relation to thematic monotonicity. For example, if a reanalysis violates constituent structure monotonicity but not thematic monotonicity, is this predicted to be dispreferred in relation to a reanalysis which does not violate monotonicity at either level? Given that the class of possible reanalyses allowed by the constituent structure based model of chapter 4 is a proper subset of that allowed by thematic monotonicity, it is natural to assume that this is so. In an implementation, this might be captured by applying the *tree-lowering* operation discussed in chapter 4 before attempting other reanalysis operations.

Finally, one aspect of our approach which we regard as an advantage over many alternative models is that our constraints are precise enough to be computationally implemented, allowing search strategies and reanalysis preferences to be tested rigorously. We will discuss some implementational details in the next chapter.

Chapter 6

Further Computational Issues

6.1 Introduction

In the previous chapter, we argued that the constituent structure based model of Gorrell (1995b) is too restrictive to capture the range of reanalyses available to the human parser.[1] We therefore proposed a new monotonic model based on thematic structure, which, we argued, has a more satisfactory range of predictions. This model, and monotonicity constraints in general, can be seen as a declarative description of the range of *cures* available to the parser. However, as we argued in chapter 4, a complete reanalysis theory must also consider procedural issues; how the parser diagnoses its errors, and the preferences through which cure is implemented. We believe that computational modelling is essential for a proper investigation of these issues, but computational models can be used in a number of different ways. The implementation which we described in chapter 4 was "tailor made" to capture the constraints of a particular reanalysis model, that of Gorrell (1995b). This served to highlight some points in which the original theoretical model was unclear, and also yielded some interesting new predictions. However, as a long term strategy, we believe that there is also a good case for building an implementation which is *general* to as great an extent as possible, so that, for example, we can easily examine and compare the consequences of adopting various proposed constraints on diagnosis and repair. In this chapter, we consider issues related to the creation of such a general parser.

We begin by describing a general method for fully connected incremental parsing, which

[1]Some of the material reported in section 6.2 of this chapter will appear as Lombardo and Sturt (1997).

we see as a prerequisite for building psycholinguistic models. This includes a discussion of the attachment operations used by the parser, and a method for dealing with non-lexical structure building and infinite local ambiguity.

After this, we move on to a discussion of the problems involved in implementing general monotonic models. In particular, we see how the reanalysis search space can be limited with reference to general monotonicity constraints based on properties. This requires the implementation of a delay mechanism, so that we can model, for example, the head-driven confirmation of thematic nodes introduced in chapter 5.

Partly as a consequence of its greater degree of generality, it will be seen that the implementation which we discuss in this chapter embodies a less restrictive view of monotonicity than that presented in chapter 4. In contrast to that model, where monotonic behaviour was imposed by a limited set of parsing operations, the implementation described here requires an architecture in which reanalysis solutions can be initially proposed by a search mechanism, and subsequently filtered out with reference to the relevant monotonicity constraints.

6.2 Fully Connected Parsing

This section describes a general method for tackling the problem of fully connected incremental parsing. As we argued in chapter 3, full incrementality is a desirable property of psycholinguistic models, and is assumed, either explicitly or implicitly, by many authors (c.f. the Left-to-Right constraint of Frazier and Rayner (1988)). However, fully connected parsing presents certain computational problems, which have to be addressed by any model which makes such an assumption. The two salient problems concern non-lexical structure building, which affects all fully incremental parsers based on lexicalized grammars, and infinite local ambiguity, which affects all parsers, regardless of the type of grammar used. In this section, we will introduce a parser which deals with both of these issues in a general manner, using methods which could be relatively simply applied to other parsers, in contrast to the rather specialized parser described in chapter 4. We will begin by discussing the problem of non-lexical structure building, and take this opportunity to introduce the main algorithms and data structures of the parser.

6.2.1 Tree Representation

Before we move on to discuss the parser in detail, we will make a brief comment on the data structures which we use to represent trees. The model described in chapter 4 used a description, consisting of a set of dominance and precedence relations between nodes, as its primary data structure. In contrast, the implementation described here will use a standard tree, encoding direct dominance and sisterhood relations. This tree can then be enriched with the information required to implement the monotonicity constraint under consideration, as we will see in section 6.3. This allows us to maintain a basic parser which is fully incremental, while also making it possible to alter and experiment with different monotonicity constraints.

6.2.2 Projection Trees and Lexicon

The grammar is expressed as a set of projection trees, each indexed by lexical category. The Projection trees include information about arguments that are taken in general by the lexical category in question. For example, the following are projection trees for the verb category and noun category respectively:

```
     S            NP
[left:[NP]]   [left:[D]]
     |            |
     VP           N'
     |            |
     V            N
```

Notice that the S node is marked with the selection feature "left:[NP]", indicating that it must take an NP daughter to the left of the head projection, and the NP node on the noun projection tree is similarly marked with "left:[D]", showing that it takes a determiner daughter to the left of the head projection.[2]

A (slightly simplified) lexical entry for a typical transitive verb will be as follows:

(lex:"eat", cat:V, right:[NP])

The NP complement is included in the lexical entry because its selection is an idiosyncratic property of this particular verb. The lexical entry can be combined with the verbal projection tree by adding the selection feature to the projection tree's VP node, licensing the VP to take an NP as a right daughter.

[2] We omit other features, such as oblgatoriness, for clarity.

In addition to the lexicon and the projection trees, the grammar includes a specification of which categories can modify one another as adjuncts. For example, the following shows that a PP can be adjoined to the right of an N′.

```
modifies(pp,nb,right).
```

6.2.3 Parsing Operations

It will be recalled that the implementation described in chapter 4 employed three basic parsing operations, *Right Attachment*, *Left Attachment*, and *Tree Lowering*. More generally, these can be characterized in terms of two basic modes of combining the projection of the new word (which we will refer to as the "projection") with the Current Partial Phrase Marker (henceforth CPPM).

1. The Projection is added as the daughter of a node on the right frontier of the CPPM.

2. The CPPM is added as the daughter of a node on the left frontier of the Projection.

Right and Left Attachment can be seen as primitive examples of cases 1 and 2 respectively. Tree-lowering may be seen as a complex operation involving both of the two primitive operations, since although the Projection is added as a daughter on the right frontier of the CPPM (case 1), it is also the case that part of the CPPM is added as a daughter on the left frontier of the projection (case 2).

Moving to a more abstract level, we can describe all the parser's actions in terms of combining two structures; the *Left Tree* and the *Right Tree*. In the parser described in chapter 4, the *Left Tree* was always instantiated by the CPPM, and the *Right Tree* by the projection of the new word. However, in the parser described in the present chapter, this will no longer necessarily be the case, making it convenient to define the two basic modes of combination in more abstract terms, as follows:

Right Combination: The *Right Tree* is added as the daughter of an accessible node on the *Left Tree*.

Left Combination: The *Left Tree* is added as the daughter of an accessible node on the *Right Tree*.

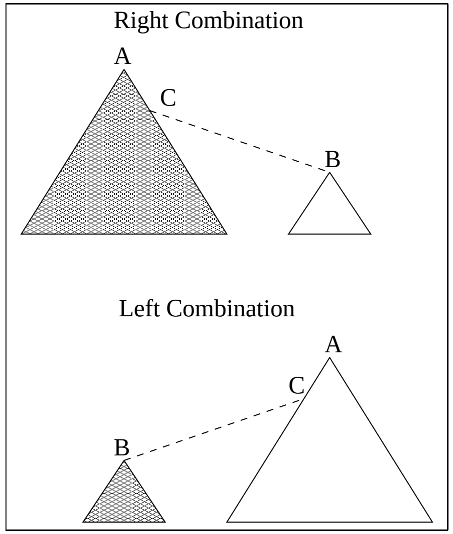


Figure 6.1:

As in the algorithm described in chapter 4, the notion of *accessibility* is used to ensure that all attachments obey the theory of trees. A node which is "accessible on the *Left Tree*" can be assumed to dominate the *Left Tree*'s right-most terminal node, and a node which is "accessible on the *Right Tree*" can be assumed to dominate the *Right Tree*'s left-most terminal node. Additionally, an accessible node on the *Left Tree* may not dominate either a node which obligatorily requires an argument or a node which is hypothesised. In the general case, an accessible node on the *Left Tree* will be on the *Left Tree*'s right frontier[3]. Conversely an accessible node on the *Right Tree* will generally be on the *Right Tree*'s left frontier[4].

These modes of combination are illustrated informally in figure 6.1. In the parser described in this chapter, we will divide parsing operations into those which are used for *adding* dependencies (i.e. attachment and adjunction), and reanalysis operations, which involve *breaking* dependencies (these include *tree-lowering* and other operations which we will describe later).

[3] A node N is on the right frontier of a tree T iff there is no node which is preceded by N in T.
[4] A node N is on the left frontier of a tree T iff there is no node which precedes N in T.

The attachment and adjunction operations are described below:

Right Combination:

Right Attachment: The *Right Tree* is added a subcategorized daughter of a an accessible node in the *Left Tree*.

Right Adjunction: The *Right Tree* is adjoined as a modifier at an accessible node in the *Left Tree*.

Left Combination:

Left Attachment: The *Left Tree* is added as a subcategorized daughter of an accessible node in the *Right Tree*.

Left Adjunction: The *Left Tree* is adjoined as a modifier at an accessible node in the *Right Tree*.

6.2.4 Attachment Behaviour

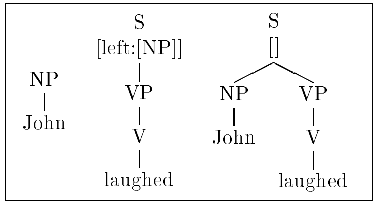
The task of a fully connected incremental parser is to take each word as it is encountered in a left to right parse, and connect it to the CPPM being built. In the parser which we describe in this chapter, we decided to take a slightly more bottom-up approach than the model given in chapter 4. Recall that in chapter 4, we employed top-down projection of hypothesised complements. So, for example, in parsing the sentence *John resents Mary*, an NP would be built at the point where *resents* is encountered. We did this in order to maintain similarity with Gorrell's model (Gorrell, 1995b), which employs this feature. However, as we mentioned in chapter 4, this approach runs into problems in cases of optional arguments, where the presence or absence of an argument cannot be predicted in advance. If a node is predicted, but is subsequently not justified by the input, then it will have to be removed, involving the breaking of at least a domination relation. It is more economical to build optional complement nodes only in response to the input. However, once this step has been taken, it seems natural to use the same approach for obligatory complements as well. In this implementation we therefore employ a general strategy of not building structure unless it serves to connect the current word with the CPPM. We will see in the following discussion that this means that hypothesised structure is only built in cases where there is no direct licensing relation between the word currently being processed and the current CPPM. Furthermore, the future addition of long distance dependencies to the grammar may well require a greater use of predictive structure building, (see Crocker (1994)).

We will now give a short illustration of each of the basic parsing operations.

Left Attachment

(6.1) John laughed.

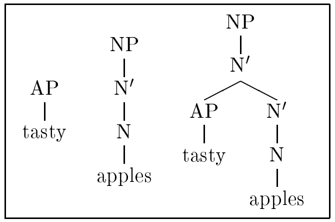
The word *John* is initially projected to an NP, and subsequently this is attached as a left daughter of the projection of *laughed*.



Left Adjunction

(6.2) tasty apples.

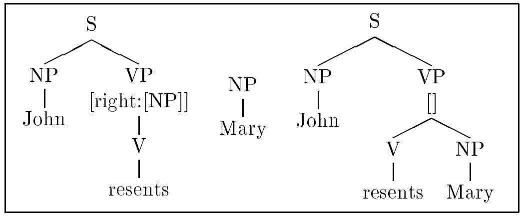
The adjective *tasty* projects to an AP, and adjoins to the N′ of the projection of *apples* (recall that information about modifiers and modifiees is recorded in the grammar):



Right Attachment

(6.3) John resents Mary.

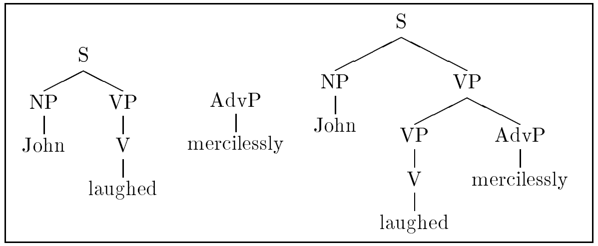
The NP *Mary* is projected to an NP, and attached as a subcategorized daughter of the VP.



Right Adjunction

(6.4) John laughed mercilessly

The adverb *mercilessly* is projected to an AdvP, and adjoined to the VP.



Ambiguity resolution strategies can be defined in cases where more than one operation can be applied, or where the same operation can be applied at more than one site. At present, the following strategies have been implemented. All except the last are adopted fairly directly from the model described in chapter 4:

- Prefer standard combination to reanalysis.
- Prefer Attachment to adjunction.
- Prefer Right Combination to Left Combination.

- Prefer combination to recently built structure.
- Prefer obligatory attachment to optional attachment.

These could be extended relatively simply to capture certain other preferences, such as a preference to attach to a verbal projection over other categories (Abney, 1989).

6.2.5 Projection Algorithm

In general, any parser constrained by full connectedness will be required to build structure which, assuming standard notions of constituency and phrase structure, is not directly licensed by lexical input. This is the case in the example discussed in chapter 4:

(6.5) Mary thinks John .....

Here, on the assumption that *thinks* subcategorizes for a clausal complement but not an NP, a clausal node has to be built which is not yet justified by a lexical head.

There are two ways in which to treat such phenomena in a parser that uses a lexicalized grammar, corresponding to the on-line and off-line computation of lexically unlicensed structure respectively. In the first of these, which we will employ here, the parser computes non-lexical structure on-line, as and when it needs it. In the second approach, a grammar is expanded off-line to include structures that are larger than the argument domain of a single lexical item. The first of these approaches represents are more transparent grammar parser relation, since it does not require grammar compilation. These two approaches are described in more detail in Lombardo and Sturt (1997). For examples of the second approach, see Lombardo and Sturt (1997) and Thompson, Dixon, and Lamping (1991).

In chapter 4, we considered the problem of non-lexical structure building, but did not give a general approach for solving it. In fact, we effectively employed both the on-line and the off-line approaches; the on-line approach is exemplified by the "non-lexical structure building" routines of the parser, while the off-line approach is implicit in the fact that there was no theoretical limit to the depth of a grammar entry's tree fragment, potentially allowing structures to cover a wider domain than the lexical anchor's local argument domain.[5].

The task of the projection algorithm is to find a path of nodes connecting each new word in the string with the current partial phrase marker. We will call such a path a *projection*

[5]see Frank (1990), (1992) for a discussion of initial trees as domains of locality in TAG

path. The algorithm works by non-deterministically searching through grammatically possible paths, by hypothesising structure that has not yet been directly justified by the input. A pseudo-code summary of the algorithm is as follows. The *Path* variable is used to keep track of the search and will be explained below.

1. Given a word W, Find PT, projection tree of W, and initiate *Path* as Ø. Go to 2.

2.
   - **If** root(PT) can combine with the Current Partial Phrase marker CPPM through right or left combination,
   - **Then** update the CPPM accordingly, and exit.
   - **Else** set *Path* to *Path* ∪ root(PT), and go to 3.

3.
   - **If** there is a projection tree PT2, to which PT can left-combine, such that root(PT2) ∉ *Path*,
   - **then** go to 2, with PT2 as PT.
   - **Else** fail.

It can be seen that the projection algorithm attempts to combine each new word to the CPPM by projecting to successively higher and higher levels of structure. The *Path* variable builds up a list of categories whose combination with the CPPM has already been attempted. Each time the projection is built up to a new root node, a check is made, to ensure that this root node does not already appear on the path. As we will see in section 6.2.6, this guards against non-termination in the case of left-recursive structures, by ensuring that the projection path is no longer than what is necessary to guarantee connectedness. If a structure is genuinely left recursive, then new nodes will be inserted into the path in response to subsequent lexical input, via the *tree lowering* operation.

We will now see an example of the Projection algorithm working with the following input sentence:

(6.6) John thinks his sister hates him.

Assume that after *thinks* has been processed, the CPPM is as shown in figure 6.2: Assume further, that the projection tree of *his* projects to the D(eterminer) category. Now, following step 1 of the the projection algorithm, an attempt is made to combine D directly with the

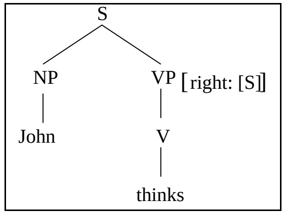


Figure 6.2:

CPPM. However, this fails, because there is no node on the right frontier of the CPPM which selects for D, or to which D can adjoin (right combination) , and furthermore, D does not select for, and cannot adjoin to the root category (S) of the CPPM (left combination). Following step 2 of the algorithm, the parser finds the Noun projection tree, which can left combine with D. However, the root of this projection (NP) still cannot combine with the CPPM, so the parser searches for a projection tree which can left-combine with NP. The result is the verbal projection tree, rooted in S, which can also right-combine with the CPPM. After the combination has taken place, the CPPM includes two hypothesised projection trees, which have yet to receive their lexical heads. It should be noted that circumstances such as these represent the only occasions in which the present parser builds structure which does not dominate lexical material. Hypothesised structure is built only when necessary to maintain full connectedness, though, as we mentioned above, the addition of long distance dependencies would increase the predictive behaviour of the parser (Crocker, 1994).

This resulting CPPM is shown in figure 6.3, where the structure built by each stage of the projection algorithm is indicated with dotted loops. If the next word is *sister*, then this will be recognized as a noun, and added as the lexical head of the noun projection, via an operation which we call *Head Attachment*, which unifies a lexical head with a predicted dangling node.

Note that the CPPM depicted in figure 6.3 is one in which a special case of left combination may apply. For example, if the next word is *little* (as in *John thinks his little sister hates him.*), then this adjective can adjoin via left combination to the hypothesised N′ node. In this case of left combination, the *Left Tree* is the projection of the new word, and the *Right Tree* is a hypothesised subtree of the new projection. The result of this adjunction is given

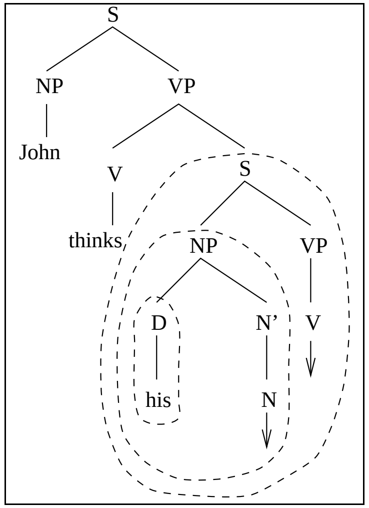


Figure 6.3:

in figure 6.4.

It can be seen that the algorithm guarantees full connectedness for a wide range of cases. However, given the grammar which we are assuming, there is a class of cases in which the algorithm, as stated above, does not allow fully connected parsing. These are cases which are exemplified in figure 6.5 Here, the new word W projects to a YP, but YP is an argument that is not selected by the X^0 projection as a general property, but rather is an idiosyncratic argument of the word which subsequently fills the lexical head X^0. In such a case, the attachment of YP is not licensed by any information available on the projection tree (which only includes non-idiosyncratic information). Situations such as these are often encountered in head final languages, such as Japanese.

In the Japanese parser described in 4, this problem did not arise, because of some grammatical assumptions which we made there. In that chapter, we used an approach in which all categories which are potentially arguments of a verbal projection (for example, case marked NPs, or clauses with complementizers) may "select" the verbal projection (effectively adjoining to it). However, as we pointed out there, this solution does not have the advantage of

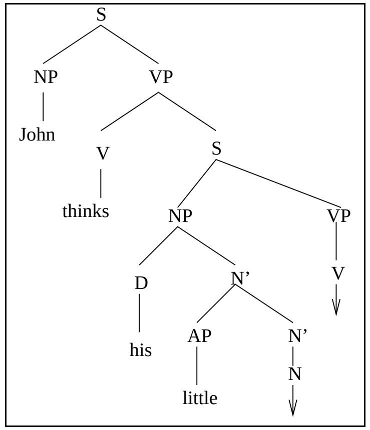


Figure 6.4:

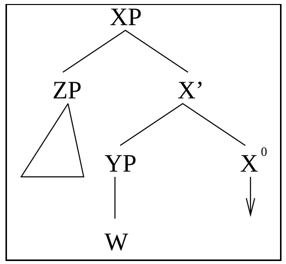


Figure 6.5:

generality, since the grammatical assumptions on which the approach relies do not necessarily hold of all languages in which this problem arises.

Of course this problem could be solved immediately if the set of projection trees were expanded to cover all possible subcategorization patterns. Pre-head structuring would then simply be a matter of finding a projection tree compatible with the arguments found at any particular parse state. However, assuming the existence of a large number of possible subcategorization patterns in a realistic grammar, this approach entails that the parser will have to cope with massive local ambiguity at the start of a phrase.

One possible approach, which would avoid these problems, is to allow some form of underspecification of subcategorization patterns, allowing the parser to narrow down its choice of subcategorization structure as the lexical information becomes available. This is possible if we define a set of projection trees in terms of an inheritance hierarchy based on possible subcategorization patterns. This approach allows the parser to hypothesise a head projection before it is justified by lexical input, and to narrow down the set of possible subcategorization patterns as each pre-verbal argument is read, moving down the inheritance hierarchy from more general to more specific levels. Assuming that it is possible to describe subcategegorization patterns in a simple hierarchical form, this approach will allow fully general connectivity in parsing head final structures, while avoiding the need to search the entire list of lexical entries for possible subcategorization patterns. An example of a parser which uses this approach is that described in Konieczny (1996) for parsing a subset of German. Although we do not develop such an approach in this thesis, we believe that it would be relatively straightforward to extend the parser in this way.

6.2.6 The Problem of Infinite Local Ambiguity

One problem that any fully connected parser has to address is that of infinite local ambiguity. Consider the following sentence fragment.

(6.7) John hates his......

Given the information available at this point in the parse, it is not possible to predict the length of the connection path between *his* and the CPPM, and, indeed, this connection path could be arbitrarily long, since *his* could mark the start of a an arbitrarily deep left recursive structure:

(6.8) a. John hates [$_{NP}$ his sister].

b. John hates [$_{NP}$ [$_{NP}$ his sister]'s boyfriend].

c. John hates [$_{NP}$ [$_{NP}$ [$_{NP}$ his sister]'s boyfriend]'s mother.]

d. John hates [$_{NP}$ [$_{NP}$ [$_{NP}$ [$_{NP_i}$ his sister]'s boyfriend]'s mother]'s cat.]

e. .... etc. *ad nauseam*

Recall that the projection algorithm checks that each new root node added to the connection path has not already occurred previously on the path (step 3 of the projection algorithm, page 145). This prevents any termination problems in the processing of left recursive structures, by ensuring that each repetition of the left recursive structure is only built in response to overt lexical evidence. The model uses the *tree lowering* operation in the processing of left recursive structures. The definition of the tree lowering operation used in this parser is similar to that defined in chapter 4. In terms of the model described here, the tree lowering operation can be seen as a complex case of right combination, which is composed of primitive combination operations.

> **Tree Lowering:** A node on the right frontier of the *Left Tree* is detached, and *left combined* with the right tree. The right tree is *right combined* with the left tree.

Let us see how the processor acts when faced with a left recursive structure such as (6.9):

(6.9) John hates [$_{NP_k}$[$_{NP_j}$ [$_{NP_i}$ his sister]'s boyfriend]'s mother.]

On reaching the word *his*, the processor initially projects this word as a determiner, and extends the projection to NP_i, which is attached as an argument of *hates*. Subsequently, the noun *sister* is attached in the predicted head position, by *head attachment* (see 6.2.6a).

(6.10) John hates [$_{NP_i}$ his sister].

Next, the processor encounters the possessive particle *'s*[6] We assume that the possessive particle projects to a determiner via the following projection tree:

D
[left:[NP]]
|
poss
|
's

[6] For simplicity, we assume that the possessive particle is processed as a separate word.

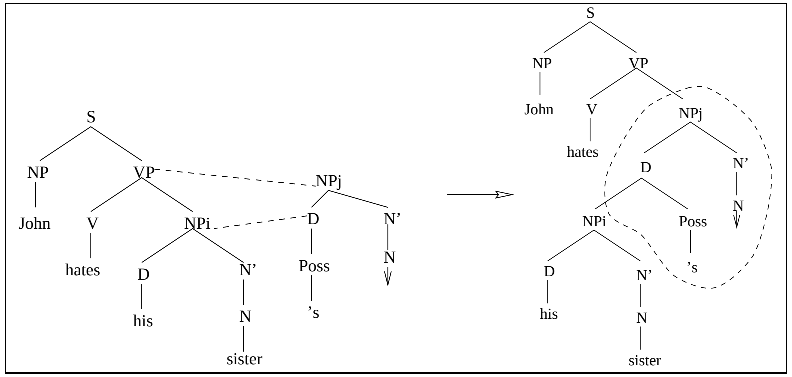


Figure 6.6: Attachment of possessive particle via tree-lowering. The dotted loop encloses the structure inserted into the CPPM.

This determiner cannot be directly combined with the CPPM, so its projection is extended to the NP level (this is possible because a determiner can *left combine* with a nominal projection tree). Now, the resulting NP (which we will call NP_j) can be combined with the CPPM via tree lowering. This is because there exists an accessible node on the right frontier of the CPPM (i.e. NP_i) which can be left-combined with the projection tree of the possessive (it is selected in the possessive's projection tree), and, moreover, NP_j is selected by a node on the right frontier of the CPPM. The way in which the possessive particle is incorporated into the CPPM is illustrated in figure 6.6. After the possessive particle has been incorporated, the noun *boyfriend* is added as the predicted head noun, via head attachment, completing the NP_j. The following possessive particle is added in the same way, this time lowering NP_j. Left recursive structure of arbitrary depth can be built in this way, through the repeated application of the tree-lowering operation.

From this discussion it can be seen that the algorithm embodies a notion of "minimal effort" in the processing of left recursive structures. Intuitively, when the grammar permits an unbounded chain of left descendant nodes, the resulting structure can be seen as a repetition of some specific pattern, which we may call a "Minimal Recursive Structure" (Lombardo and Sturt, 1997). The projection algorithm ensures that no more than one link in the chain of Minimal Recursive Structures is built at one time, and, moreover, the general bottom-up

character of the parser ensures that each link is only built in response to lexical input.

6.3 Monotonicity and Indexing

So far, we have said little about reanalysis, and almost nothing about how reanalysis algorithms can be constrained by monotonicity. Perhaps the most economical way to implement a monotonicity constraint is to employ the relevant structural relations directly as the main data structure of the parser, and to use a limited set of parsing operations and search procedures which jointly guarantee monotonicity, so that the parser is incapable of even proposing reanalysis solutions which would violate the monotonicity constraint under consideration. This, effectively, is how we implemented the model described in chapter 4. However, it cannot be guaranteed that such solutions can be found in the general case, where the model might be less restrictive. In this section, we will describe a general method of representing the required structural relations. Given this representation, we will be able to determine, given a possible detachment node, and a possible target site for reattachment of this node, whether this reattachment obeys the monotonicity constraint as defined by some generator property P. This allows one to determine the consequences of defining monotonicity with respect to any arbitrary generator property. In order to be employed in an algorithm, however, the technique also needs to be supplemented by search procedures which allow the processor to *propose* candidate detachment and target sites. The representational device which we use is an adapted version of the indexing technique originally proposed by Latecki (1991) for implementing command relations.

6.3.1 Generalized Command Theory and Indexing

By way of introduction, we will describe how the indexing technique used by Latecki (1991) can be used to define command relations, before discussing how it can be applied to the definition of reanalysis constraints. We will start by giving a summary of the notion of *generalized command*, as defined by Barker and Pullum (1990), though for the formal definitions, the reader is referred to the original paper. First, we will illustrate how the notion of *generator property* is used in generalized command theory. Barker and Pullum (1990) noticed that most command relations in the syntactic literature have the following form:

X *P*-commands *Y* iff for every node *N*, such that *N* has property *P*, and *N* properly dominates *X*, then *N* dominates *Y*.

The required command relation can then be obtained by specifying the value of the predicate *P*. For example, one version of the well-known *c-command* relation can be given by specifying *P* as the property of being a *branching node* (i.e. a node which has more than one daughter), while the notion of *m-command* used in Barriers (Chomsky, 1986a) can be defined by specifying *P* as the property of being a maximal projection.

In what follows, we will show how Latecki's (1991) indexing technique allows one to calculate, given two nodes, *X* and *Y*, whether the *P*-command relation holds between *X* and *Y*. In the indexing scheme, all nodes in the tree are labelled with the following information:

1. An index, which encodes whether or not the generator property holds of the node in question. We will refer to the index of node *N* with respect to the generator property *P* as $Id(N, P)$, or simply as "the index of *N*", as appropriate.

2. A set of indices inherited from the node's ancestors. We will refer to the set of indices for node *N* with respect to property *P* as $Set(N, P)$, or "the index set of *N*", as appropriate.

Now, for each node *N*, the indices are assigned and inherited in the following way:

1. If the property *P* holds of *N*, then the value of $Id(N, P)$ is a singleton set containing a unique integer. Otherwise, the value is the empty set.

2. If *N* is the root node, then $Set(N, P) = Id(N, P)$. Otherwise, the value is $Set(M, P) \cup Id(N, P)$, where *M* is the mother node of *N*.

To illustrate the indexing scheme, assume that we want to implement a command relation in which the relevant generator property is that of being a branching node. Then, assuming a tree as given in figure 6.7. It can be seen that each node is labelled with its *Id* (above) and its *Set* (below). It can be seen each node which holds the generator property (i.e. each branching node) has a singleton set containing a unique integer as its *Id*, and the *Set* of each node is the union of that node's *Id* and its mother's *Set*. Latecki (1991) shows that, given such a scheme, it is possible to determine whether a command relation holds between any two nodes by making a simple subset check. Specifically, the following statement is true:

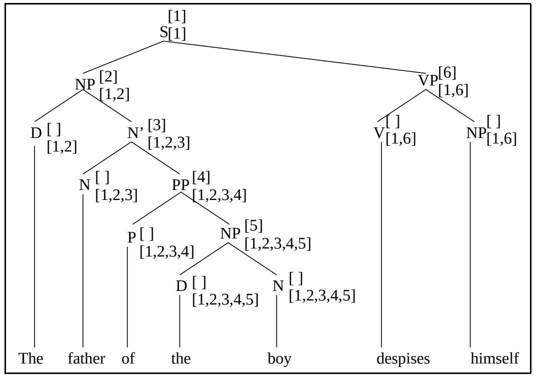


Figure 6.7: Illustration of Latecki's indexing technique. Each node is labelled with its *Id* (above) and its *Set* (below), with *"branchingness"* as the relevant generator property.

X *P*-commands *Y* iff $(Set(X, P) - Id(X, P)) \subseteq Set(Y, P)$.

As an example, let us say that we have c-command defined as a command relation with "branchingness" as the relevant generator property, and that c-command is one of the conditions that has to hold for an antecedent to bind its anaphor. We can then determine that the NP headed by *father* c-commands the NP *himself*, because $([1, 2] - [2]) \subseteq [1, 6]$. On the other hand, the NP headed by *boy* does *not* c-command the NP *himself*, because $([1, 2, 3, 4, 5] - [5]) \not\subseteq [1, 6]$.

6.3.2 Generalized Monotonicity and Indexing

In this section, we will see how the indexing method described above can be used for the implementation of generalized monotonicity constraints based on properties. We will begin with a discussion of the *P*-dominator monotonicity constraint, which we repeat below for convenience:

P-Dominators:

The *P-dominators* for *Y* at state *S* (abbreviated as $D(Y, P)^S$) is the set of all nodes *X*, such that property *P* holds of *X* at state *S*, and *X* properly dominates *Y* at state *S*.

P-Dominator Monotonicity:

For each transition $S \Longrightarrow T$, and for each node X, $D(X,P)^S \subseteq D(X,P)^T$.

Assume that we have the same index assignment scheme as Latecki (1991) uses for command relations, and that the indexing information for any node is available at each state.

For any node N, the set of indices corresponding to N's P-dominators is simply $Set(N,P)$ minus $Id(N,P)$. Now, in order to ensure that P-Dominator monotonicity is maintained, we must ensure that indices are only ever *added* to a node's index set, and not removed.

During reanalysis, if reattachment of some node N is attempted, we can perform a check to ensure that the P-dominators of N form a subset of the P-dominators of the candidate target site. This can be achieved by checking the following condition on reattachment:

***P*-Dominator Condition on Reattachment:** The reattachment of N as a daughter of M obeys P-Dominator Monotonicity iff $(Set(N,P)$ minus $Id(N,P)) \subseteq Set(M,P)$.

P-Dominator monotonicity can therefore be seen as a command condition across state transitions. That is, for any transition $S \Longrightarrow T$, if the position of some node N changes, then the position of N at state S must P-command the position of N at state T.

As an illustration, consider a monotonicity constraint defined in terms of the generator property of being a *thematic assigner* node (i.e. the extended maximal projection of a thematic assigner). Now consider the (translated) relative clause ambiguity from De Vincenzi and Job (1995) illustrated in figure 6.8: It can be seen that all thematic assigner nodes are given an *Id* consisting of a singleton set containing a unique integer, and the indices are inherited through the tree as before. Recall that De Vincenzi and Job's experiment showed evidence that the relative clause was initially attached low. Given this initial attachment, then, it can be determined that reanalysis to the high site (indicated by the dotted line) is possible, because $([1,2,3]-[3]) \subseteq [1,2]$. In contrast, consider the condition in which a thematic domain boundary intervenes between the two attachment sites, given in figure 6.9: Here, it can be seen that reattaching the relative clause to the high site will violate monotonicity as defined in terms of thematic assignerhood, since $([1,2,3,4]-[4]) \not\subseteq [1,2]$.

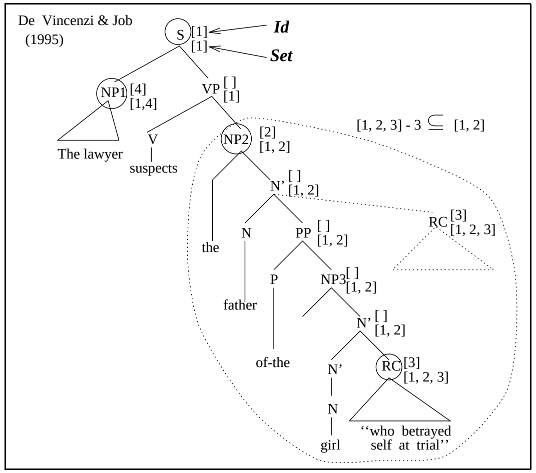


Figure 6.8:

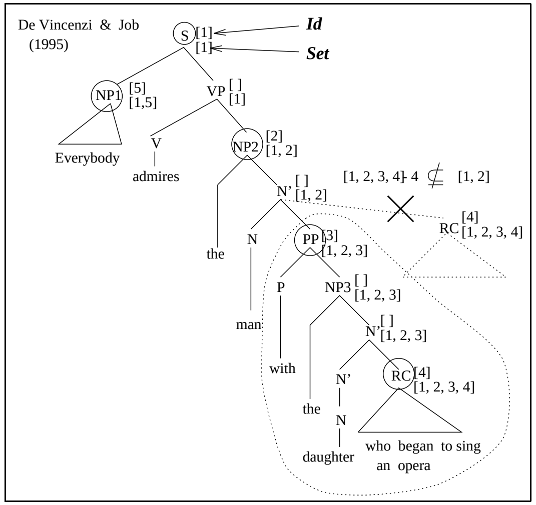


Figure 6.9:

Checking P-Precedence Partner Monotonicity

We now turn to a consideration of how the conditions on precedence relations can be checked using Latecki's indexing system. In general, the problem of checking the coherence of precedence relations is more complex than that of checking the coherence of dominance relations, and at first sight, it seems that the Latecki indexing technique is inappropriate for checking precedence relations, since the propagation of indices through the tree follows exclusively *domination* paths. However, we will see that, given certain assumptions about the processing algorithm used, the indexing technique can indeed be useful for this purpose.

Recall that P-Precedence Partner Monotonicity requires that, for any two nodes X and Y, which both have property P, once a precedence relation between X and Y has been determined, it cannot be "revoked". The definition is as follows:

P-Precedence Partners:

The set of *Precedence Partners* at state S (written $PRP(P)^S$) is the set of all ordered pairs of nodes X and Y, such that property P holds of both X and Y at state S, and X precedes Y at state S.

P-Precedence Partner Monotonicity:

For each transition $S \Longrightarrow T$, $PRP(P)^S \subseteq PRP(P)^T$

There are three ways in which a precedence relation between X and Y can be revoked in the transition from state S to state T:

1. X precedes Y at state S, but Y precedes X at state T.

2. X precedes Y at state S, but X dominates Y at state T.

3. X precedes Y at state S, but Y dominates X at state T.

These three cases represent transitions that we wish to rule out if we want to preserve P-Precedence Partner Monotonicity. The first case is illustrated in figure 6.10. As far as this case is concerned, we cannot rule out this exchange of the precedence relation between two nodes using just the Latecki Indexing technique, since a transition involving only this re-ordering will result in no change in the indices or index sets associated with each node. In figure 6.10, both X and Y have the generator property at state S, so they each have an index.

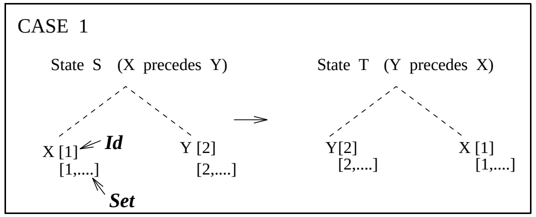


Figure 6.10:

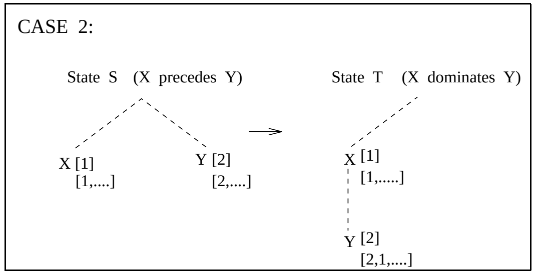


Figure 6.11:

But the re-ordering has no effect on the indices. We would have to rule out this case through some other constraint in the algorithm. For example, if we make the reasonable assumption that the precedence ordering of the actual terminal nodes of the string cannot be changed (c.f. Weinberg (1993)), and assume a parsing algorithm which is bottom-up to the extent that every node of which P holds is guaranteed to dominate at least one terminal node, then "flipping" will always result in a non-monotonic update to the index set of some node.

Cases 2 and 3 are illustrated in figures 6.11 and 6.12 respectively. The fact that both of these are non-monotonic follows from the exclusivity condition of the theory of trees, which says that no two nodes may be both in the dominance and the precedence relation. In both cases it can be seen that either one of X or Y inherits an index into its *Set* which had already been assigned at state S. In general, if we assume an incremental parser which maintains a totally connected structure at all states, and that indices are systematically inherited in the manner described above, then we can rule out cases 2 and 3 by ensuring that any index

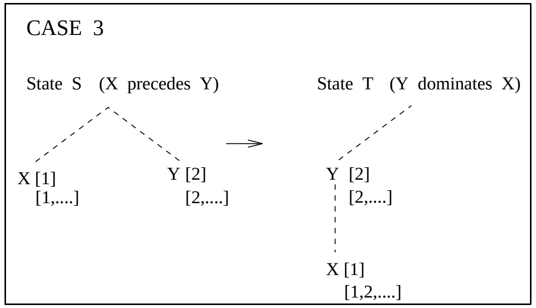


Figure 6.12:

added, in the transition $S \Longrightarrow T$, to the index set of a node that has property P is a *newly assigned* index (i.e. the index did not exist at state S). Recalling figure 6.11, the addition of the index 1 to $Set(Y,P)$ is ruled out because this index already existed at state S. Similar remarks hold for the addition of index 2 to $Set(X,P)$ in figure 6.12.

The reader can therefore determine that, if a node N does not dominate any node which has the generator property P, then the P-Dominator condition on reattachment (defined above) suffices to guarantee monotonicity of both Dominance and Precedence constraints for this generator property. On the other hand, if N *does* dominate some node which has the generator property P[7], then, in order to guarantee P-precedence partner monotonicity in the reattachment of N as a daughter of M, we need to apply not only the P-Dominator condition on reattachment, but also check that each item which is a member of the index set of M, but which is not a member of the index set of N is a newly assigned index at the current state.

The check as it stands seems rather complex. However, it can be simplified. Adding a new index (i.e. assigning an *Id* to a node where none previously existed) during a state transition will never result in a violation of monotonicity, as this will only result in the creation of *new* dominators and precedence partners. Therefore, we can adopt the convention that, during each state transition, new indices are only added *after* all attachment and reanalysis operations have been performed. This means that we do not need to consider the newly assigned indices when checking for P-precedence partner monotonicity, and the following check suffices:

[7] Note that this includes the situation in which N itself has the property P.

P-path nodes: N is a P-path node iff N dominates some node which has property P.

P-Monotonicity Condition on Reattachment:

- If N is not a P-path node, then the reattachment of N as a daughter of M obeys P-monotonicity iff $(Set(N,P) - Id(N,P)) \subseteq Set(M,P)$.
- If N is a P-path node, then the reattachment of N as a daughter of M obeys P-monotonicity iff $(Set(N,P) - Id(N,P)) = Set(M,P)$.

Before the discussion proceeds, we should point out that, although the check for whether a node N is a P-path node would appear to require an unbounded search of the entire subtree dominated by N, this is not in fact, the case. In the current implementation, at the point where a node N is assigned the generator property P, a feature is added to N, marking it explicitly as a P-path node, and the same feature is added to all nodes on the domination path between N and the nearest P-path node dominating N. This means that, to check whether or not a node is a P-path node, it is sufficient to check for the presence or absence of a local feature on that node.

We will now see an example of this constraint in action. Assume that we have a monotonic theory in which the relevant generator property is that of being a thematic node (i.e. either a thematic assigner or a thematic receiver). Recall the sentence *The boy put the candy on the table into his mouth.*. Assuming an initial attachment of the PP *on the table* as the daughter of VP, then the reanalysis of this PP as a modifier of *candy* violates the monotonicity constraint. In figure 6.13, we see how this can be determined using the indexing technique. Since the PP is a Path node (i.e. it dominates at least one node of which the generator property holds), its P-dominators need to be identical to those of the target site indicated by the dotted line. However, as can be seen, this is not the case. If the PP were *not* a path node, by contrast, then the reanalysis would be possible, since the P-dominators of the PP form a subset of those of the target site.

6.4 Dynamic Index Creation and Propagation

So far, we have been discussing Latecki's (1991) indexing technique in *static* terms. However, for the purposes which we have in mind here, we need to be able to add indices during the

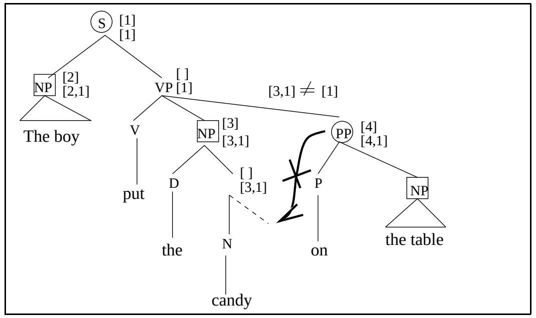


Figure 6.13:

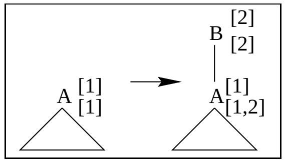


Figure 6.14:

course of a parse. If we were to confine ourselves to a purely top-down algorithm, without delay in the assignment of indices, this would be a trivial problem, because each node would create its *Set* simply by appending its *Id* to the set of its mother. However, in the present model, not only are we dealing with an algorithm that is not purely top-down, but also we allow the assignment of an index to be delayed with respect to the creation of that node. We will take these two problems in turn.

6.4.1 Index Propagation

Consider the problem of bottom-up structure building first. If a node B is added as the mother of a node A, as in figure 6.14, then the contents of B's *Set* must be added, not only to A's *Set*, but to the *Set* of every node in the subtree dominated by A. As long as structure building is guaranteed to be purely bottom-up, it is possible to propagate the indices in the required manner by representing the *Set*s as lists with variable tails, and using unification

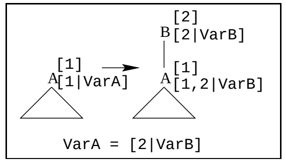


Figure 6.15:

for the adding of indices from mother to daughter. Taking figure 6.14 again as an example, A's initial *Set* will be represented as `[1|VarA]`,[8] B's *Set* will be represented as `[2|VarB]`, and the required propagation of indices will be achieved through the unification equation `VarA = [2|VarB]`.[9] This is illustrated in figure 6.15. If nodes are later added over B, then the new index can be propagated through the subtree dominated by A through unification with `VarB`. The reader will verify that if this scheme is maintained, unification can be used as the mechanism through which propagation occurs.

However, this propagation scheme is not satisfactory either. To see why this is so, consider the case in which a node C is inserted between nodes B and A. Let us say that the *Id* of C is [3], and therefore, we want to add 3 to the *Set* of A and to those of all nodes dominated by A. If this is done through unification with `VarB`, then the new index will also be added to B's *Set*, which goes against the required downward propagation of indices. This is illustrated in figure 6.16 Our solution to this problem is to use *residuation* (coroutining), in such a way that the tail variable of a mother's *Set* is never fully unified with that of the daughter's *Set*. Let us say again that the tail variable of A's *Set* is `VarA`, and a new index 2 is being added from B using the list term `[2|VarB]`. This time, instead of unifying `VarA` with `[2|VarB]`, as we did in the previous case, we unify `VarA` with `[2|VarC]`, where `VarC` is a new variable. We then set this new variable `VarC` to wait for `VarB` to become further instantiated using what we call a *residuation link*.[10] Let us say a new node C is added as the mother of B, with index 3. This index is added to B's *Set* by unifying B's tail variable with `[3|VarF]` (where `VarF` is

[8] In this chapter, we will use PROLOG syntax, where lists are delimited by square brackets, the vertical bar "`|`" separates the head from the tail of the list, and variables start with upper case characters. See any introductory PROLOG text for more details

[9] Again, we are using prolog conventions, where "`=`" is to be interpreted as unification, rather than equality.

[10] This is implemented using functional residuation in the LIFE programming language.

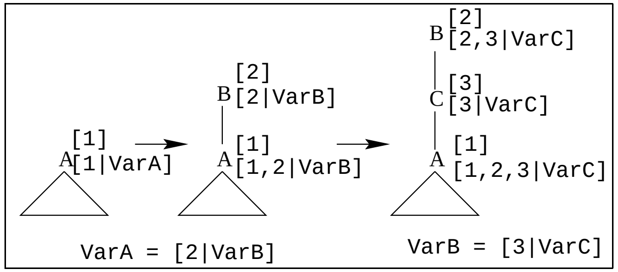


Figure 6.16:

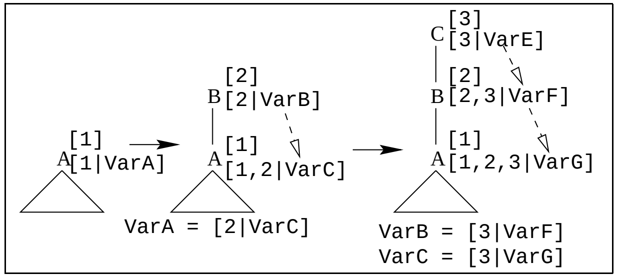


Figure 6.17: Illustration of Coroutining implementation of index propagation. A dotted arrow pointing from X to Y indicates a *residuation link*; that is, variable Y is waiting for variable X to become instantiated.

a new tail variable). Now that B's tail variable is instantiated, A's tail variable is triggered to inherit the new index, and the same mechanism will trigger any other nodes dominated by A to inherit the new index in the same way. This is illustrated in figure 6.17.

The core of the inheritance mechanism is implemented via the following LIFE function:

```
inherit([X|Y],Var) -> tail_var(Var) = [X|NewVar], inherit(Y,NewVar).
```

The first argument of `inherit` contains the tail variable of the mother's *Set*, and the second argument contains the tail variable of the daughter's *Set*. The function does not fire until the mother's tail variable is instantiated to a list that is at least as specific as `[X|Y]`.[11] When it is

[11] In LIFE, the evaluation of a function is delayed until the calling term entails the definition term.

instantiated in this way, the daughter's tail variable is returned by the function `tail_var`[12], and instantiated to `[X|NewVar]`, where `X` is the index being inherited, and `NewVar` is the new variable. The function is then called with the mother's new tail variable as its first argument, and the daughter's new tail variable as its second argument. Again, the function will not fire until the mother's tail variable becomes further instantiated, at which point, any new indices added to the mother's *Set* will be inherited by the daughter's *Set*. In cases where a mother-daughter relation is broken during parsing, for example, where one node is inserted between two existing nodes, or a node is "moved" from one position to another in the tree, we need also to break the corresponding residuation link between the mother and daughter, and assign a new link between the daughter and its new mother. This is necessary to prevent the daughter from subsequently inheriting spurious indices in its new position, and also to allow it to inherit any required indices from its new mother.[13]

The advantage of this index inheritance mechanism is that propagation of indices is guaranteed to be purely top-down, regardless of the direction in which structure is built.

For example recall the situation mentioned above, in which C is added, not as the mother of B, but inserted between B and A. Then, the residuation link between `VarB` and `VarC` is broken, and new residuation links are added, between B and C, on the one hand, and C and A on the other, as shown in figure 6.18. This allows the correct propagation of indices, in contrast to the situation shown in figure 6.16.

6.5 Delayed Assignment of Indices

In this section, we return from a general discussion of the indexing technique to the *specific* problem of implementing the thematic monotoniocity model.

Recall from the discussion of the model in chapter 5 that thematic assigners and receivers are confirmed according to a head-driven scheme:

Thematic Confirmation:

Two nodes A and R are *confirmed* as Thematic Assigner and Thematic Receiver respectively iff

[12] New indices might have been added to `Var` since the function was first called.

[13] The "breaking" of a residuation link is achieved simply by matching a non-recursive clause of the `inherit` function, thereby terminating the daughter's wait to inherit the mother's next index.

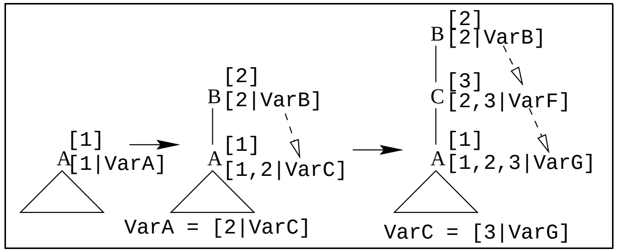


Figure 6.18: Illustration of Coroutining implementation of index propagation. A dotted arrow pointing from X to Y indicates a *residuation link*; that is, variable Y is waiting for variable X to become instantiated.

i. R is in a position from which it may receive a thematic role from the head of A.

ii. The head words of both A and R are instantiated.

iii. The resulting thematic dependency is semantically acceptable.

In what follows, we will discuss the implementation of this delay mechanism. For the purposes of the implementation, we will ignore condition iii, because we do not have a representation of the world knowledge necessary to check semantic acceptability.

The implementation of the head-driven index assignment is again achieved through the use of residuation (or coroutining). To begin with, we will give a more detailed description of the representation of nodes in the system.

Every node N includes the following information:

- One set, *Id*, which contains a unique integer if N is confirmed as a thematic *assigner*, and is empty otherwise (this will be called N's *A-Id*).
- One set, *Set* representing thematic assigners dominating N (this will be called N's *A-Set*).
- One set, *Id*, which is a set containing a unique integer if N is confirmed as a thematic *receiver* (this will be called N's *R-Id*).
- One set,*Set*, representing thematic receivers (this will be called the node's *R-Set*).

As we saw above, projection trees and lexical items are marked with selection features which show the categories of possible arguments. These can be either *thematic* or *non-thematic*.

In case of a *thematic* selection feature licensing the attachment of some node N as a daughter of M, an index assigning function is called. This function waits for the heads of both N and M to be instantiated by the input, and then *confirms* the thematic dependency as follows:

1. If M's maximal projection is not already confirmed, then a unique index is assigned to the *A-Id* of M's maximal projection, and added to the *A-Set* of M's maximal projection.
2. A unique index $I2$ is assigned to N's *R-Id*, and added to R's *R-Set*.

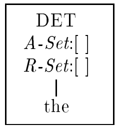
The indices are added to the *A-Set* and *R-Set* in the form of open-ended lists, as outlined above, and when each index is assigned, it is propagated to the *Sets* of all descendants as well.

We will now give an example parse to see the indexing technique in action. Consider the following input sentence:

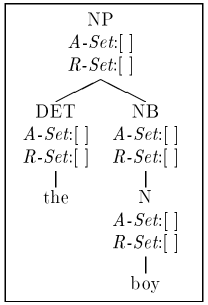
(6.11) The boy put the candy on the table.

In the following diagrams, we will display the *A-Set* and *R-Set* of each node. Each node's *A-Id* or *R-Id* will be represented by underlining the appropriate index in the appropriate *Set* of that node.

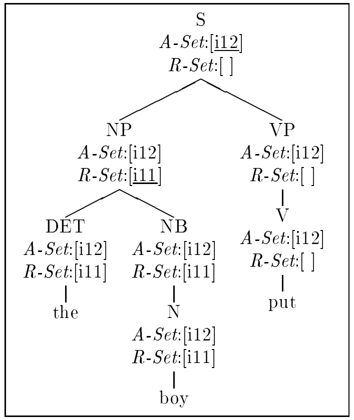
The parse begins with the input of the word *the*, which is projected to a determiner category. Since it is not yet dominated by either thematic assigners or receiver nodes, its *A-Set* and *R-Set* are both empty.



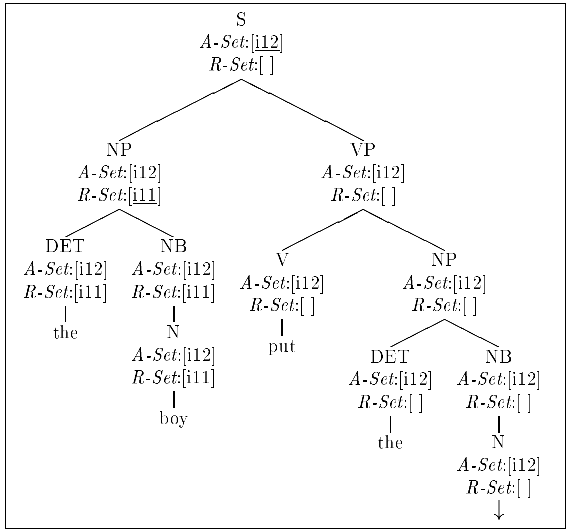
When the next word *boy* is read, it is projected to an NP, and the determiner is *left-combined* to this node:



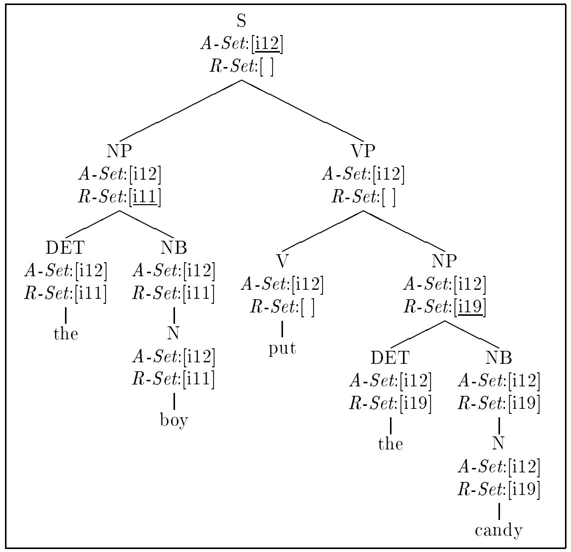
Next, the verb *put* is read and projected to an S node. The NP is *left-combined* with this projection as its subject. Since the subject is specified to receive a themtic role, and the heads of both projections *boy* and *put* are instantiated, indices are added to the relevant maximal projections; a receiver index to the NP and an assigner index to the S. These indices are inherited appropriately:



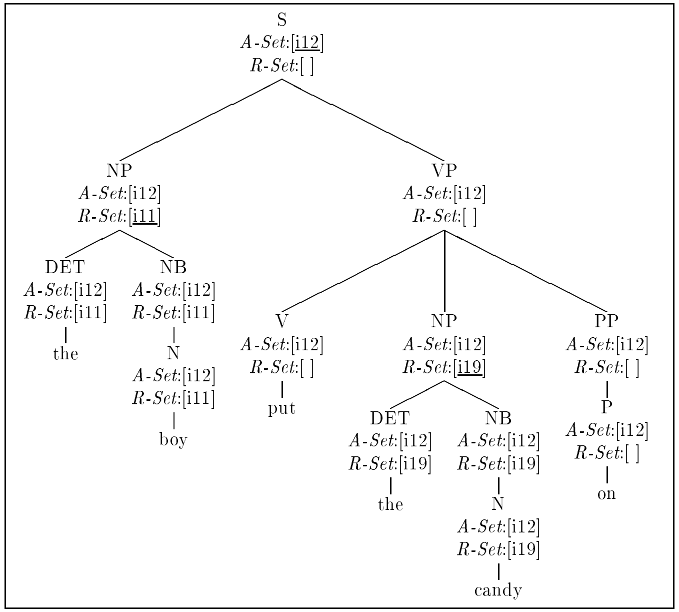
The word *the* is encountered. It is projected to the NP level (using the projection algorithm outlined above), and attached as a complement of *put*. The dependency is specified as thematic, but since the head of the NP has not yet been read, its *R-Id* is delayed:



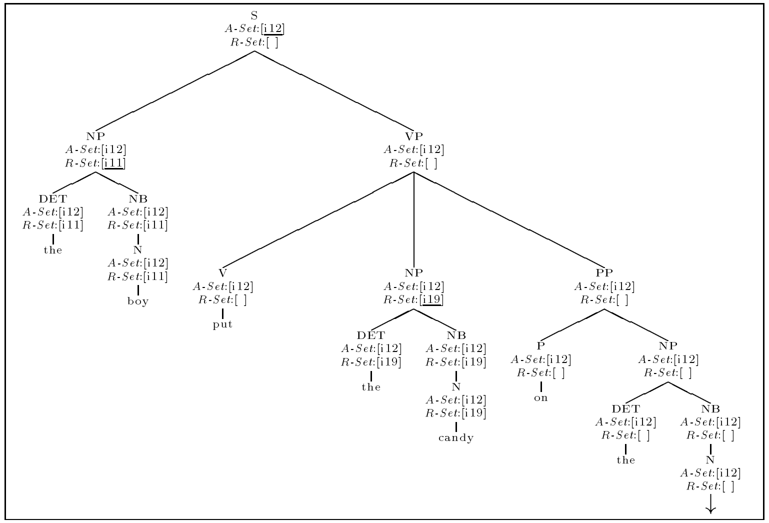
When *candy* is read, it is attached as the head of the NP, triggering the assignment of its receiver index:



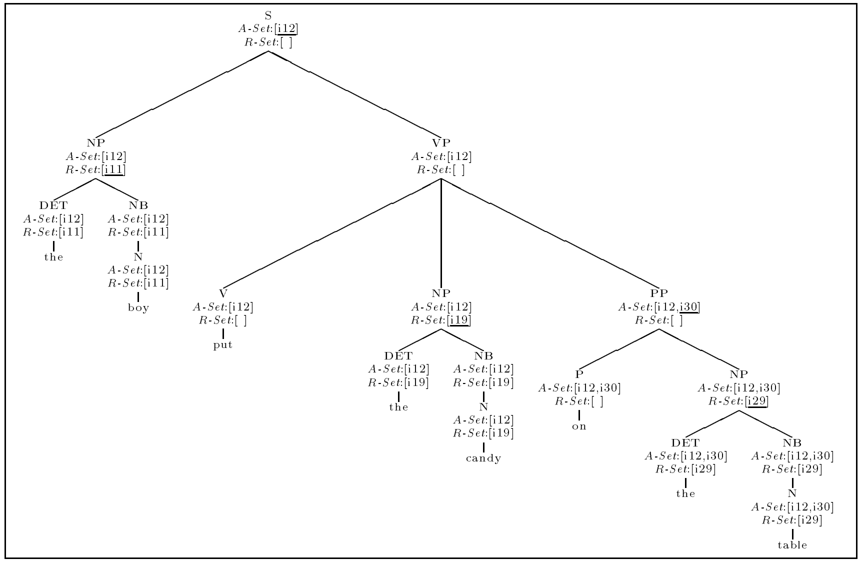
Next, *on* is projected to a *PP*, which is attached as the second complement of *put*. We assume that the LOCATIVE argument of *put* is an “indirect” argument, in which the *preposition* assigns the thematic role (cf. chapter 5). Therefore, the PP will be a thematic assigner node when the head of the following NP is instantiated. However, as yet, no *A-Id* is assigned:



The determiner *the* is read, and projected to the level of NP. This NP is attached as the complement of the preposition:



Finally, the noun *table* is read and attached as the head noun of the NP. Since both the heads of the PP and the NP are now read, an *A-Id* is assigned to the PP, and an *R-Id* to the NP. These indices are then inherited as appropriate:



6.5.1 Indexing as a Constraint on Reanalysis

As an illustration of how the head-driven assignment of indices can be used to limit search during reanalysis, consider the case of the PP in the above example, before and after *table*, the head of its NP complement, is read. Let us assume that we have a reanalysis operation which detaches a node N from its current position, and *right combines* N with an accessible node to its left. Let us say, further, that the notion of accessibility relevant to this reanalysis operation is *left accessibility*. Intuitively a node M is *left accessible* to N if M is to N's "immediate left". The notion can be defined as follows:

> A node M is *left accessible* to N iff M precedes N, and there is no node L such that M precedes L and L precedes N.

Before the noun *table* is read, we have the configuration displayed in the penultimate diagram for the example parse above. Assume that we are dealing with a monotonicity constraint which is defined in terms of the generator property of being a *thematic node* (i.e. a node that is confirmed as either a thematic assigner or a thematic receiver node). The set of nodes

which are *left accessible* to PP consist of the following:

⟨ N, NB, NP ⟩

From the definition on page 160, we can determine that the PP node is not a P-path node, because (in the penultimate diagram) it does not dominate any nodes which have the generator property. This means that we can determine whether the reattachment of PP at site X satisfies monotonicity by performing the following check:

$(A\text{-}Set(\text{PP}) - A\text{-}Id(\text{PP})) \subseteq A\text{-}Set(X)$

and

$(R\text{-}Set(\text{PP}) - R\text{-}Id(\text{PP})) \subseteq R\text{-}Set(X)$

The reader can verify that, before the word *table* is attached, the PP can be *right combined*, with any of its *left accessible nodes*, since PP's *A-Set* (i.e. [i12]) is a subset of the *A-Set*s of all these nodes, and the same is true for PP's *R-Set* (i.e. the empty set). In particular, the grammar allows the PP to be adjoined to the N′ node as a modifier of *candy*. However, now consider the situation after the noun *table* has been attached, and the assignment and receipt of the locative role has been confirmed. This is illustrated in the final diagram of the example parse. This time, PP *is* a P-path node, because it dominates a node that has the generator property (in fact it dominates two; both itself and its daughter NP). This means, according to the definition given on page 160, that the following conditions have to be satisfied in order to allow the reattachment of PP as the daughter of X:

$(A\text{-}Set(\text{PP}) - A\text{-}Id(\text{PP})) = A\text{-}Set(X)$

and

$(R\text{-}Set(\text{PP}) - R\text{-}Id(\text{PP})) = R\text{-}Set(X)$

Now, it can be seen that all of the PP's *left accessible nodes* are blocked. This means, for example, that the possibility of reattaching PP as a modifier of *candy* is lost after the head noun *table* has been attached, predicting difficulty involved with the continuation in the following sentence:

(6.12) #The boy put the candy on the table into his mouth.

As we discussed in chapter 5, if we allow semantic plausibility as a condition on the confirmation of thematic nodes, then the following contrasts are also predicted:

(6.13) a. #Mary saw the man with the binoculars through the telescope.

b. Mary saw the man with the moustache.

(6.14) a. #The politician put the report on the desk into his briefcase.

b. The politician put the report on the riot into his briefcase.

6.6 Checking Relational Monotonicity Constraints

So far we have seen how Latecki's indexing technique might be applied to implement monotonicity constraints based on *Properties*. However, we have seen in chapter 5, that thematic monotonicity is defined in terms of *Visibility*, which is a *relation*, treating modifiers in a manner which is distinct from non-modifiers. In this section, we will briefly consider how such relational constraints can be implemented using the indexing scheme described above. Although a discussion of a *general* scheme for implementing *relational* constraints would lead us too far afield, we will sketch how the relation of *visibility* can be represented within the indexing scheme proposed here. It will be recalled that this relation defines all *thematic assigner* nodes to be visible to *modifiers*, and all *thematic nodes* (i.e. both assigners and receivers) to be visible to non-modifiers, with the exception that all nodes *inside* a modifier inherit the modifier's "blindness" to thematic receiver nodes that are *outside* the modifier. Stated in these terms, the reader could be forgiven for thinking this definition to be contrived and awkward. However, it is much more natural when we consider how it would be implemented in terms of assignment and inheritance of indices. In a monotonic system with *visibility* as the generator relation, we label each node with two *Id*s and *Set*s, corresponding to *thematic receivers* and *thematic assigners*, as we have done in the examples above. The *thematic assigner* indices are assigned and inherited from mother to daughter just as though *thematic assignerhood* were a generator property. However, by definition, *thematic receiver* indices would never be assigned to modifiers, and, furthermore, we make the assumption, modifiers only *inherit* thematic *assigner* sets, and not thematic *receiver* sets. Thus no indices of thematic receivers dominating a modifier would be inherited by that modifier's descendants. The reader can verify that if

this scheme is followed, the R-dominators with respect to the visibility relation can be readily determined for any node.

6.7 Concluding Remarks

In this chapter, we have considered some general questions concerning the implementation of incremental, monotonic parsers. The implementation extends the model described in chapter 4 by creating a more general and complete treatment of fully connected parsing, and provides a useful basis for the further exploration of issues related to reanalysis. At the time of writing, the implementation is still very much in a prototype stage; for example, although some basic reanalysis mechanisms exist (for example, tree-lowering, and searches of left accessible nodes) we do not have the full set of reanalysis search mechanisms necessary to implement all the examples which we used to motivate the thematic monotonicity model in chapter 5. However, we hope to have convinced the reader that the prototype implementation which does exist can be extended in the future to investigate psycholinguistic issues further. In particular, we foresee the implementation of the following items:

- A full *diagnosis* component, which will be able to propose optimal search strategies in response to the error signal, and whose search mechanisms can be constrained independently of *cure* search (see Lombardo (to appear) for an example of such a parser which incorporates such a component).
- The addition of greater linguistic coverage, by incorporating long distance dependencies, passivisation, raising and control into the grammar.
- The addition of a mechanism to deal with linguistic category ambiguities.
- An implementation of lexical and syntactic statistical preferences.

Chapter 7

Reanalysis Ambiguity: An Experimental Study

7.1 Introduction

This chapter reports two experiments designed to test for reanalysis preferences. Specifically, the experiments test the hypothesis that there exists a recency preference in reanalysis, such that, all other things being equal, and under certain structural conditions, the processor prefers to revise a more recent than a less recent decision. We will examine structures similar to the example we discussed in chapter 4, and in Sturt and Crocker (1996), which we repeat below in (7.1):

(7.1) a. I know the man who believes the countess hates herself.

b. I know the man who believes the countess hates himself.

In (7.1), assuming that the NPs headed by *man* and *countess* are each attached as NP objects to their respective immediately preceding verbs, then the appearance of *hates* will force reanalysis. However, there are two possibilities for this reanalysis, depending on which of the two possible NP attachments is revised. In chapter 4 and in Sturt and Crocker (1996), we describe the *tree-lowering* operation, which is applied to perform revisions of this nature as well as the attachment of postmodifiers and certain types of conjunctions. We hypothesised there that the preferences for the application of this parsing operation would be similar to those for the attachment of postmodifiers, in other words, that we would find a *recency*

preference in reanalysis of this type, similar to the late closure effect found for modifiers in English, and we modelled this by formulating a bottom-up search procedure for *tree-lowering*. This would predict that (7.1.a) would be preferred over (7.1.b), as the gender marking on the reflexive pronoun *herself* is consistent with *countess*, while that on *himself* is not. This recency preference is also predicted by other accounts. For example, it follows as a result of the decay of activation in Stevenson's (1994b) competitive activation model, as well as from the principles governing the *Attach Anyway* operation in the diagnostic based model of Fodor and Inoue (1994), (to appear). On the other hand, a recency-based reanalysis preference is not the only logical possibility. For example, it might very well be that reanalysis procedures are sensitive to some variety of semantic or discourse accessibility that favours higher nodes over lower nodes, perhaps because their semantic content is more directly linked to the main content of the utterance as a whole (see, for example, the *Relativized Relevance* principle of Frazier (1990b)).

Intuitively, (7.1.a) does seem to be easier to process than (7.1.b), thus providing initial support for the recency hypothesis. However, this intuition needs to be backed up by experimental evidence before we can claim that the preference is generalizable to new items and new subjects. This brings us to the question of which experimental techniques are appropriate for testing reanalysis preferences. There have been very few experiments examining reanalysis for its own sake, and so this is still very much an open question. One technique which has been used successfully is the Speeded Grammaticality Judgement task, with Rapid Serial Visual Presentation (Warner and Glass (1987), Ferreira and Henderson (1991)), in which the subject has to read each sentence displayed rapidly, one word at a time, and decide whether or not it is grammatical. In this task, processing difficulty is indicated by increased numbers of grammatical sentences being judged as ungrammatical, presumably because the increased pressure imposed by the rapid serial presentation causes subjects to lose track of the parse in cases where cognitive resources are in heavy demand. However, rapid serial presentation is very far removed from normal reading, and may encourage subjects to adopt unnatural strategies. This seems to be a particular danger when one is examining reanalysis processes, which are known to rely on regressive eye-movements to previous portions of the sentence (Frazier and Rayner, 1982). A technique which does not allow regressions may therefore encourage subjects to reanalyse in an unusual way. This is a potential problem in any experimental procedure where previous portions of the stimulus disappear, or become obscured by other characters, as the subject reads the stimulus, as is the case, for example, in non-cumulative self-paced reading, and rapid serial visual presentation. In cases where the experiment is designed to demonstrate the presence versus the absence of reanalysis, such as standard experiments investigating initial attachment preferences, this does not matter so much. However, where it is the reanalysis processes themselves that are under examination, as in the experiments discussed in this chapter, it seems preferable to use procedures which pose the smallest possible disruption to normal reanalysis processes. Perhaps the ideal technique is eye-tracking, since this combines extremely fine-grained data resolution with something approaching a normal reading situation. However, as the experiments reported here are very much preliminary investigations, it was decided to use simpler and less labour-intensive techniques. The first experiment uses cumulative self-paced reading, and the second uses a simple off-line questionnaire.

7.2 Experiment 1

This experiment was a self-paced reading study, which was designed as an on-line test of reanalysis preference. The experiment involved items which each contained two verbs that could subcategorize either for a tensed clause or an NP. The materials were constructed in such a way as to force readers to reanalyse, while at the same time leaving two possibilities for revision; to perform reanalysis at a low (recent) site, or alternatively at a high (less recent) site. The revision itself was later disambiguated using a manipulation involving a reflexive/reciprocal anaphor (e.g. *herself*, *each other*). Two control conditions were added, in which a complementizer *that* unambiguously signalled one of the two complements as a tensed clause[1]. An example item in the four conditions is given in (7.2):

[1] In fact, in the "Unambiguous High" condition, there remains a local ambiguity (see the discussion in section 7.4), but the results of experiment 1 shows that, the complementizer acted as an effective disambiguator.

(7.2) a. **Ambiguous[2] Low:**
The judge accepted the witness who recognised the suspects had been deceiving themselves.

b. **Ambiguous High:**
The judge accepted the witness who recognised the suspects had been deceiving himself.

c. **Unambiguous Low:**
The judge accepted the witness who recognised that the suspects had been deceiving themselves.

d. **Unambiguous High:**
The judge accepted that the witness who recognised the suspects had been deceiving himself.

It was assumed that in the two ambiguous conditions, readers would initially attach both of the two NPs *the witness ...* and *the suspects* as direct objects of the respective verbs, causing a need to reanalyse at the point where *had* appears. Thus, we would expect to find some processing disruption in the two ambiguous conditions compared with the two unambiguous conditions. However, depending on the preference for reanalysis, we would also expect to find a difference between the two *ambiguous* conditions. That is, if the processor adopts a low (i.e. recent) reanalysis preference, we would expect a greater disruption in the Ambiguous High condition than in the Ambiguous Low condition. On the other hand, if the processor adopts a high reanalysis preference, we would expect to find the opposite effect.

The methodology used here can be seen as an extension of the standard garden-path technique developed originally by Frazier and colleagues (Frazier, 1978; Frazier and Rayner, 1982) which we discussed in chapter 2, section 2.3. In the garden-path technique, attachment preferences are tested by presenting subjects with a local ambiguity, and subsequently disambiguating this ambiguity downstream. In the case of the experiment reported here, we present the subject with *two* local attachment ambiguities, and subsequently present the processor with a *reanalysis* ambiguity, allowing either one, but not both, of the previous ambiguous attachment decisions to be revised. We then disambiguate the reanalysis ambiguity

[2] The term "ambiguous" is here being used as a shorthand for "locally ambiguous".

in favour of either the preferred or the dispreferred reanalysis, as hypothesised by the theory, and expect to find greater ease of processing where the disambiguation is consistent with the (predicted) preferred reanalysis relative to the dispreferred reanalysis. We will call such an effect a *second order* garden path.

7.2.1 Method

Subjects

Twenty eight students from the University of Glasgow were paid to participate in this experiment. All subjects were native speakers of English.

Materials and Design

Thirty six experimental items were constructed in the format shown in example (7.2) above, yielding a 2×2 design (Ambiguity (unambiguous vs. ambiguous) × Attachment site (high vs. low)). A list of the materials can be found in appendix A, section A.1.

The critical verbs for the experiment were taken from the following list of nine verbs, each of which can take either an NP or a tensed clause as its complement:

reveal, accept, recognise, forget, recall, find, acknowledge, understand, see

As the main point of the experiment was to test preferences in reanalysis, we wanted to maximize the chances of an initial NP attachment of the post-verbal material in the two ambiguous conditions, thereby ensuring that reanalysis really does occur in these conditions. It has been argued (Trueswell, Tanenhaus, and Kello, 1993) that processing disruptions at the point of disambiguation in NP/S ambiguities do not necessarily reflect reanalysis processes. They claim that subcategorization frequency influences the immediate attachment of complements of these verbs, and that, in some cases, processing disruption at the point of disambiguation may simply be the result of a spill-over of a surprize effect due to the lack of a a complementizer, rather than a revision of an initial NP attachment.

In order to rule out possible confounds of this nature, we required that all verbs used in the experimental materials should be biased towards NP subcategorization. To test this, we took random samples of 100 occurrences of each experimental verb from section A of the British National Corpus (roughly 14 million words), and, for each sample, compared the

number of times that the verb immediately preceded an NP complement versus a tensed clause complement (either with or without a complementizer). For all of the nine verbs used in the experiment, the number of NP continuations was at least double the number of clausal continuations. Six of the verbs (i.e. *reveal*, *accept*, *forget*, *recall*, *understand*, and *see*) were found to be NP biased in the norming study reported in (Trueswell, Tanenhaus, and Kello (1993)). Of the remaining three verbs, two (i.e. *recognize* and *find*) were found to be NP biased in the norming study reported in Holmes, Stowe, and Cupples (1989).

To control for possible biases due to differences in the *relative* preferences of each verb, we ensured that each verb appeared in the high site (matrix clause) and the low site (embedded relative clause) an equal number of times.

Furthermore, it was ensured that each of the anaphors (*himself*, *themselves*, *itself*, *herself* and *each other*) disambiguated in favour of the high reading and the low reading an equal number of times.

A pretest was conducted to test the semantic plausibility of the two possible disambiguations of each item. The subjects had to rate the plausibility of the content of the two complement clauses that would result from the high reanalysis/attachment on the one hand, and the low reanalysis/attachment on the other. So, for example, the two pretest materials corresponding to the item given in (7.2) were as follows:

(7.3) **Low disambiguation:**
The suspects had been deceiving themselves.

(7.4) **High disambiguation:**
The witness had been deceiving himself.

Pretest booklets were constructed, each containing 72 sentences similar to the above (i.e. two for each of the 36 experimental items) in random order. Twenty one students of the University of Glasgow participated in the pretest. The subjects were asked to rate the plausibility of each sentence on a scale from 0 (very implausible) to 7 (very plausible). The results of the pretest show no evidence for a plausibility bias in the materials. A t-test taking items as the random factor, and comparing the plausibility of the high disambiguation against the plausibility of the low disambiguation for each item, showed that there was no statistical difference between the two disambiguations; $t(35) < 1$.

Procedure

The experiment employed a two-region cumulative self-paced reading paradigm. Despite the fact that non-cumulative procedures have been found to be more sensitive to on-line effects (Just, Carpenter, and Wooley, 1982; Ferreira and Henderson, 1990), the cumulative procedure was chosen because it allowed the subjects to re-read previous portions of the sentence, thereby minimizing the disruption to normal reanalysis processes.

The regions are illustrated with respect to the example item repeated below, where the # symbol represents the region boundary:

(7.5) a. **Ambiguous Low:**
The judge accepted the witness who recognised the suspects # had been deceiving themselves.

b. **Ambiguous High:**
The judge accepted the witness who recognised the suspects # had been deceiving himself.

c. **Unambiguous Low:**
The judge accepted the witness who recognised that the suspects # had been deceiving themselves.

d. **Unambiguous High:**
The judge accepted that the witness who recognised the suspects # had been deceiving himself.

The second region always began at the final tensed verb, which, in the two ambiguous conditions, signals that reanalysis must take place.

The materials were rotated according to a Latin Square design, so that each subject saw only one condition of each material. The 36 experimental materials were embedded in a list of 80 filler sentences, also divided into two regions, none of which included any of the critical experimental verbs. The sentences were presented in a random order for each subject.

The experiment was implemented on a Macintosh computer using the PsyScope experimental software package developed at Carnegie Mellon University. At the beginning of each trial, an asterisk appeared in the centre left of the screen, indicating the position of the left edge of the first word. When the subject pressed the middle button of the button box, the

	Unambiguous	Ambiguous
High	2049	2478
Low	2136	2440

Figure 7.1: Mean reading times in milliseconds, for the second region.

first region of the sentence appeared, with the second region obscured by dashes. When the subject pressed the central button for a second time, the dashes were replaced by the second region, and the words of the first region remained in place. On 16 of the experimental trials, and 26 of the filler trials, a comprehension question followed the presentation of the sentence, which required either a "yes" or a "no" answer. The subjects responded to the questions by pressing either the right (Yes) or the left (No) button on the button box, after which the asterisk for the next sentence appeared. In the experimental trials, the question directly probed the attachment of the critical region (for example, "Had the suspects been honest with themselves?"). For trials with no questions, the phrase "No question" appeared, and the subjects had to press either the left or right button to move to the asterisk for the next sentence. The 116 trials were preceded by a screen of instructions, and ten practice trials.

7.2.2 Results

Before performing statistical analysis on the data, all trials exhibiting either extremely high or extremely low reading times were excluded from all further consideration. The criterion for exclusion was calculated according to the following procedure. First, all trials that included a region with a reading time of less than 500 ms or more than 10,000 ms were removed. This excluded less than 3% of the data. Of the remainder, all trials that included a region exhibiting a reading time of more than 2.5 standard deviations either side of that subject's mean for that region were replaced by this cut-off value. Figure 7.1 shows the mean reading times in milliseconds for the critical second region. A 2 × 2 Analysis of Variance was computed on the mean reading times for the critical second region, taking subjects (F_1) and items (F_2) as random factors. The results showed a main effect of ambiguity; $F_1(1, 27) = 19.706$, $p < .001$, $F_2(1, 35) = 7.619$, $p < .01$, but no main effect of attachment site; $F_1 < 1$, $F_2 < 1$, and no significant interaction between ambiguity and attachment site; $F_1(1, 27) = 1.149$, $p < .3$, $F_2 < 1$.

	Unambiguous	Ambiguous
High	87 %	76 %
Low	86 %	92 %

Figure 7.2: Percentage of correctly answered comprehension questions for the four conditions

Analyses were also performed on comprehension accuracy, as measured by the percentage of correctly answered comprehension questions for those 16 experimental items that were followed by questions. Figure 7.2 shows the results of comprehension accuracy for the four conditions. Analyses of Variance on comprehension accuracy revealed a main effect of attachment site, that was significant by subjects only; $F_1(1, 27) = 5.421$, $p < .05$, $F_2(1, 15) = 1.394$, $p < .26$, but no main effect of ambiguity in either the subjects or the items analysis; $F_1 < 1$, $F_2(1, 15) = 1.057$, $p < .33$. However, ambiguity and attachment site interacted significantly in both analyses; $F_1(1, 27) = 5.118$, $p < .05$, $F_2(1, 15) = 8.318$, $p < .05$. Means comparisons using t-tests revealed that the difference in comprehension accuracy between the two ambiguous conditions was significant, $t_1(27) = 3.12$, $p < .01$; $t_2(15) = 3.13$, $p < .01$, while the difference between the unambiguous conditions was not (both t's < 1). This, as well as the interaction found in the comprehension accuracy data, suggests that the comprehension disadvantage for the high attachment only has an effect when reanalysis is necessary.

7.2.3 Discussion

The most striking aspect of the results from the current perspective is the divergence between the on-line and off-line data. The reading time data shows a main effect of of ambiguity, which simply reflects difficulty associated with the need for reanalysis (a standard garden path effect), but no effect of attachment site, showing no preference for either the more recent or the less recent attachment. Conversely, the comprehension accuracy data shows a main effect of attachment site, providing evidence for a recency preference, which is present only when reanalysis is necessary (i.e. a "second order" garden path effect), but no main effect of ambiguity, showing that the mere need for reanalysis does not *per se* disrupt comprehension accuracy. The comprehension accuracy data suggests that there is indeed a recency preference for the type of reanalysis under investigation here. However, if this is so, then we are left with the question of why the recency preference was not picked up in the reading time data.

There are a number of possible reasons for this, which we outline below.

First, it may be that certain aspects of the experimental design were at fault. As we have mentioned above, cumulative self-paced reading, which we used here, has been argued to be a weaker measure of on-line processing behaviour than the non-cumulative technique (Just, Carpenter, and Wooley, 1982; Ferreira and Henderson, 1990). In addition to this, it is known in eyetracking experiments that "wrap-up" effects may contribute unwanted variance to reading times in a final region, obscuring any effects that actually exist. This might have been the case in the current experiment. In the ambiguous conditions, both the initially disambiguating word (the tensed verb that shows at least one NP attachment should be reanalysed as a clause) and the globally disambiguating word (the final anaphor) appear in the final region, and are thus susceptible to being swamped by variance related to the "wrap-up" effect. However, the fact that a main effect of ambiguity was obtained shows that at least the initial disambiguation was immune to the problems posed by both the cumulative technique and the final region wrap-up effect. Another possible problem with the experimental design is that both disambiguations appeared in the same region. If the initial disambiguation had been placed in a different region from the final disambiguation, it might have been possible to separate the effects associated with each.

Secondly, it may be that comprehension accuracy is simply a better measure of certain processing effects, such as the reanalysis effects at issue here, than self-paced reading. De Vincenzi and Job (1995) display a similar difference between on-line and off-line measures in a series of experiments investigating the effect of thematic domains on reanalysing relative clause attachments.

One possible explanation of the findings of this experiment involves the notion of *internal repair*. Konieczny, Scheepers, and Hemforth (1994) have found evidence which suggests that, in certain conditions, readers do not reanalyse when faced with an inconsistency in the input, but rather perform an *internal repair*, which is a non-monotonic constraint relaxation process that allows the processor to fit the inconsistent input item into the already interpreted structure. With reference to the results reported here, the internal repair would be performed in the ambiguous high condition, where the reader would initially mis-reanalyse low, but rather than performing a re-reanalysis, would incorporate the inconsistent anaphor into a structure with which it does not agree. This would explain the drop in comprehension accuracy for the ambiguous high condition.

However, we should be wary of basing our claims on the comprehension accuracy alone. The comprehension questions were only attached to 16 of the 36 items, which were not necessarily representative of the items as a whole. Unfortunately, not all of the controls that were applied to the materials as a whole also apply to the subset of sixteen items which were accompanied by questions. For example, these sixteen items do not preserve the balance of occurrences of the same verb appearing in both high and low positions.

Furthermore, for each of the 16 items, the high and low conditions were accompanied by different comprehension questions. This was done in order to make the correct answer (yes / no) consistent across the four conditions of each item, but it could have caused a possible confound if the high questions were in general more difficult to answer than the low questions. This confound, if it exists, could explain the main effect of attachment site seen in the subjects analysis, but would not explain the interaction.

In order to check whether or not the comprehension results reflect genuine processing preferences, it was decided to carry out an off-line questionnaire as a second experiment.

7.3 Experiment 2

In the previous experiment, we found the predicted interaction between ambiguity and attachment site in the off-line measure (comprehension accuracy) but not in the on-line measure (reading time). Experiment 2 was designed to test whether there really is an off-line preference for the low reanalysis option, as suggested by the comprehension accuracy findings. In order to do this, we used a questionnaire, in which subjects had to decide on their preferred interpretation of globally ambiguous versions of the experimental materials from Experiment 1.

7.3.1 Method

Subjects

24 students of the university of Edinburgh community were paid to participate in the experiment.

Materials and Design

The materials for the experiment were constructed by creating globally ambiguous versions of the materials for Experiment 1. This was achieved by altering the number or gender features on one of the NPs in such a way that both possible reanalysis sites could be compatible with one final anaphor. In addition to this ambiguous condition, two unambiguous conditions were used, one for each possible attachment. These were constructed, as in the unambiguous conditions of experiment 1, by using complementizers. A list of the materials used for this experiment is given in appendix A, section A.2. An example item is given in 7.6:

(7.6) a. **Ambiguous:**[3]
The judge accepted the witness who recognised the suspect had been deceiving himself.

b. **Unambiguous Low:**
The judge accepted the witness who recognised that the suspect had been deceiving himself.

c. **Unambiguous High:**
The judge accepted that the witness who recognised the suspect had been deceiving himself.

Each sentence in the booklet was followed by two sentences, representing the two alternative interpretations of the sentence, and marked “a” and “b”, which the subjects had to choose as the preferred interpretation. The alternative interpretations for (7.6), for example, were be as follows:

a The witness had been deceiving himself.

b The suspect had been deceiving himself.

The two unambiguous control conditions were added to ensure that subjects did not adopt unnatural strategies, such as simply taking the most recent available antecedent for the anaphor without regard for syntactic structure. It was possible to create globally ambiguous versions of almost all the original materials by simply changing the number agreement of one of the

[3] In contrast with the previous experiment, “Ambiguous” here should be interpreted as “globally ambiguous”.

nouns. However, there were three materials for which the globally ambiguous version required a slightly greater change, due to pragmatic considerations. These three materials are shown below. We illustrate the changes for each one, showing firstly the ambiguous low condition from Experiment 1, followed by the ambiguous condition from Experiment 2.

(7.7) **Experiment 1:**

The butler saw the prince who revealed his mistress had hanged herself.

Experiment 2:

The butler saw the prince who revealed his boyfriend had hanged himself.

(7.8) **Experiment 1:**

The workers recognised the company chairman who accepted the payrise couldn't justify itself.

Experiment 2:

The workers recognised the company directors who accepted the pay increases couldn't justify themselves.

(7.9) **Experiment 1:**

The people accepted the constitution which recognised the dictator would never establish himself.

Experiment 2:

The people accepted the constitution which recognised the dictatorship would never establish itself.

Procedure

The materials were rotated according to a Latin square design so that each subject saw only one version of each item. Experimental booklets were created which included the 36 experimental sentences interspersed among a number of filler sentences. The order of presentation of the sentences was randomized for each subject, and no two experimental items appeared adjacent to each other. The subjects were simply asked to read each sentence, followed by the two possible interpretation sentences, marked by "a" and "b", and to write the letter (either "a" or "b") which corresponded to their favoured interpretation. For each condition, "a" and "b" each corresponded to low and high attachment an equal number of times.

	Unambiguous High	Unambiguous Low	Ambiguous
% of High decisions	82 %	11 %	38 %
% of Low decisions	18 %	89 %	62 %

Figure 7.3: Percentage of High and Low attachment decisions for the three conditions

7.3.2 Results

The results of the questionnaire study are given in figure 7.3.

The three conditions were each analysed independently. The null hypothesis that there is no difference between the number of high and low attachment decisions yields an expected population mean of 50% of high attachment decisions for each condition. A t-test was used to compare this figure of 50% with the actual numbers of high attachment decisions, for each subject (t_1) and each item (t_2).

Not surprisingly, the two control conditions both differed significantly from the population mean in the expected directions, indicating that the complementizers acted as effective disambiguators; Unambiguous High, $t_1(23) = 7.53$, $p < .001$, $t_2(35) = 14.35$, $p < .001$, Unambiguous Low, $t_1(23) = 16.5$, $p < .001$, $t_2(35) = 16.37$, $p < .001$.

The ambiguous condition also differed significantly from the population mean, with significantly fewer high attachment decisions than would be expected on the null hypothesis; $t_1(23) = 3.18$, $p < .01$, $t_2(35) = 2.75$, $p < .01$. This demonstrates a recency preference in this construction.

7.3.3 Discussion

These results confirm the hypothesis that there exists a recency preference in the type of reanalysis structures that we have been investigating, and lead us to conclude that the interaction seen in the comprehension accuracy data in experiment 1 reflects true processing preferences.

One point which strikes us as interesting is the difference in accuracy between the unambiguous low and unambiguous high conditions in this questionnaire. That is, the interpretation consistent with the grammatical attachment was chosen in 89% of the time in the low condition as opposed to 82% in the high condition. Quite how much we should make of this difference remains unclear, though, since the difference is significant by items only ($t_1(23) = 1.61$, $p = .122$, $t_2(35) = 2.36$, $p < .05$; results are for a t-test comparing the number of grammatically correct interpretations chosen for the unambiguous low versus unambiguous high conditions). If this reallly does reflect a genuine difference, then it could be that there is a general low preference in these structures quite independent of whether or not reanalysis is needed in the processing of the structure (see discussion in section 7.4).

One other question whose answer remains unclear from the these results is that of whether we can claim the existence of competitive effects in the resolution of high versus low reanalysis. Despite the fact that a low preference was established in the ambiguous condition, it is clear that subjects were not simply incapable of choosing the interpretation consistent with the non-recent reanalysis, since they chose this interpretation 38% of the time. It could therefore be that the site at which the reanalysis operation is performed is not chosen according to a serial search, as suggested in chapter 4, but that both sites compete concurrently, as suggested by competition-based accounts such as Stevenson (1994b). However, a questionnaire study such as that reported in this section can only tell us whether or not a preference exists. We can conclude nothing about the on-line processes underlying the preference[4]. Further studies, manipulating the relative preferences of the recent and non-recent sites (for example, manipulating subcategorization frequencies of the two verbs) might serve to shed light on this question.

7.4 General Discussion

Taken together, the two experiments reported here demonstrate a recency preference in the resolution of the reanalysis ambiguity under consideration. Let us reconsider the four conditions from Experiment 1 repeated below:

[4]Note that in a questionnaire study, subjects are likely to become aware of the ambiguity being tested, which could well encourage them to consider a dispreferred reading which may not have been noticed in normal reading.

(7.5) a. **Ambiguous Low:**
The judge accepted the witness who recognised the suspects had been deceiving themselves.

b. **Ambiguous High:**
The judge accepted the witness who recognised the suspects had been deceiving himself.

c. **Unambiguous Low:**
The judge accepted the witness who recognised that the suspects had been deceiving themselves.

d. **Unambiguous High:**
The judge accepted that the witness who recognised the suspects had been deceiving himself.

The reading time data from Experiment 1 shows clearly that the two ambiguous conditions are processed with more difficulty than the two unambiguous conditions. On standard assumptions, this indicates that the processing of these two conditions involves reanalysis. Both the comprehension accuracy data from Experiment 1 and the questionnaire data from Experiment 2 demonstrate that, when this reanalysis is possible at either the recent or the non-recent site, the recent site is preferred.

One point, which we have not mentioned up to now is that the results demonstrate what we might call a "structure preservation" preference, in addition to the recency preference. Our labelling of the last condition in (7.5) above as "Unambiguous High" is, strictly speaking, not accurate, since this condition actually includes a local ambiguity; the word *had* could be incorporated either through standard attachment, in the high site, or otherwise, through reanalysis in the low site, by taking *the suspects* as its subject, as would be the case in the following continuation:

(7.10) The judge accepted that the witness who recognized the suspects had been deceiving themselves, was going to be a key to the prosecution case.

That is, the processor is faced with a choice of either attaching to comparatively *non-recent* structure, but preserving all dependencies built up to that point (which is consistent with the continuation in the Unambiguous High condition of the experiment) or, alternatively, attaching to more *recent* structure, but breaking a dependency (consistent with the continuation in 7.10). The processing model discussed in chapter 4 predicts that the first of these alternatives is preferred, since it involves simple attachment as opposed to tree-lowering, and we designed the experimental materials under the implicit assumption that this was correct. Such a preference would be consistent with any model which includes a preference for *minimal revisions* (c.f. Frazier and Clifton (to appear)) or some similar principle. However, an alternative account might consider recency to be a stronger preference than structure preservation. For example, Fodor and Inoue (to appear) discuss the following example:

(7.11) The reporters who knew the Swedish actress hated her mother.

This example involves a very similar interplay between recency and structure preservation, in that the word *hated* could be attached either as the matrix verb, or as an embedded verb, taking *the Swedish actress* as its subject. On the assumption that *the Swedish actress* is initially attached as the NP direct object of *knew*, the attachment of *hated* as an embedded verb would involve reanalysis, while the matrix attachment would not. Fodor and Inoue's intuition is that *hated* preferentially attaches as the embedded verb, via reanalysis. They explain this in terms of *Attach Anyway*, which is a parsing operation that incorporates each word into the Current Partial Phrase Marker in the least ungrammatical way, subject to attachment preferences including recency. The result of "attaching anyway" might well be a structure which is momentarily ungrammatical, and this situation then has to be rectified via a chain of local adjustments in the tree. In cases such as (7.11), Fodor and Inoue claim that recency wins out over grammaticality.

Extrapolating this claim to our own experimental materials, the *Attach Anyway* model would predict that in the Unambiguous High condition in (7.5), the word *had* would initially be attached, via reanalysis, to the most recent site, taking *the suspects* as its subject. If this action were taken, we would expect to find a garden path effect in the Unambiguous High condition as compared with the Unambiguous Low condition, since the local application of "attach anyway" is incompatible with the globally grammatical structure of the Unambiguous High condition. However, there is no strong evidence for such a garden path effect. No significant differences are observable between the unambiguous high and unambiguous low conditions in either the reading time or the comprehension accuracy data in experiment 1. In

experiment 2, while we saw a difference in comprehension accuracy between the unambiguous high and unambiguous low conditions, this was significant only in the items analysis, and the preference for the grammatical reading of the unambiguous high condition was still very highly significant. This pattern of data would be unexpected if the processor were applying "attach anyway" at a recent site, while ignoring the possibility of making a grammatical attachment at a less recent site. Instead, the results indicate that, at least in the construction that we are investigating here, structure preservation wins out over recency.

However, there is a difference between (7.11) and the unambiguous high condition of (7.5), which may be relevant. This contrast is more clearly brought out if we make a more controlled comparison, as in (7.12) below, where (7.12.a) corresponds to the unambiguous high condition of (7.5), and (7.12.b) corresponds to (7.11):

(7.12) a. The judge accepted that the witness who recognised the suspects had been deceiving himself.

b. The witness who recognized the suspects had been deceiving himself.

The non-local attachment of *had* in (7.12.a) involves attachment into a site which has been licensed by previous input, given the assumption that the complementizer *that* licenses an INFL projection (see Crocker (1996) for an example of a model which makes such an assumption). However, in (7.12.b), *had* has no such licensed structure into which it can attach, and so (7.12.b) could be predicted to be more difficult than (7.12.a). Notice also that, on the assumption that the NP headed by *witness* is initially posited as the root of the Current Partial Phrase Marker in (7.12.b), the non-local attachment of *had* will involve creating a new root (the matrix clause), and subordinating the NP as a daughter of this node. Removing a node from its temporarily posited position as the root of the CPPM is predicted to involve a certain amount of difficulty in some models; for example, in the competitive activation model of Stevenson (1994b), it would involve breaking the activation link between an abstract "stack node" and the CPPM root. In such a model, assuming that this difficulty is enough to "tip the balance" between recency and structure preservation, we would expect different attachment behaviour of *had* in the two sentences given in (7.12). It would also be relatively straightforward to incorporate this pattern of attachment into the model described in chapter 4 here. Technically, this would involve adding an abstract stack node similar to Stevenson's, which always dominates the root of the CPPM. In this case, "Left Attachment" would be just a special case of the tree-lowering operation, where the previous root of the CPPM is "lowered" under a new root, which is itself inserted under the stack node. However, our own intuitions on structures such as (7.12.b) and (7.11) are far from clear, and experimentation is needed to confirm these claims.

Chapter 8

Reanalysis and Left Branching Attachment: An Experimental Study

In this chapter,[1] we discuss two questionnaire studies which investigate how thematic structure affects the final interpretation of relative clause ambiguities in Japanese. We will argue that the effects found in these experiments should be interpreted in terms of constraints on reanalysis.

8.1 Introduction

In previous chapters we have discussed examples in which relative clauses can attach to either one of two possible sites in a complex noun phrase, as in (8.1), from Cuetos and Mitchell (1988):

(8.1) Somebody shot the servant of the actress who was on the balcony.

In head final languages, where recursive structure is predominantly left branching, the relative clause *precedes* both of the nouns which it may eventually modify. This can be seen in the Japanese translation of (8.1), given in (8.2):

[1]The questionnaire reported as Experiment 1 in this chapter was presented as Branigan, Sturt, and Matsumoto-Sturt (1996).

(8.2) barukonii ni iru joyuu no mesitukai wo dareka ga utta.
balcony LOC is actress GEN servant ACC somebody NOM shot.

This is in contrast to predominantly right branching languages such as English or Italian, where both possible attachment sites have been read in the input at the point where the relative clause is attached. It seems natural that a maximally incremental processor will attempt to attach the relative clause to the first site that becomes available, (i.e. the low attachment site headed by *joyuu* ("actress") in example (8.2)), and, as we discussed in chapter 5, Kamide and Mitchell (1996) found evidence for this strategy in their self-paced reading study. However, recall that Kamide and Mitchell (1996) also found evidence for an off-line preference for the high attachment of the relative clause, in a questionnaire study, that is, an interpretation in which the relative clause modifies *mesitukai*. This suggests that the eventual high attachment preference is derived via a reanalysis from low- to high-attachment.

8.1.1 Attachment Assumptions

Before we continue the discussion, we will outline our assumptions regarding the on-line attachment behaviour of the parser in the type of complex NP given in (8.2), and mention some possible alternatives to these.

Immediate Attachment

We assume firstly that the parser immediately attaches the relative clause to the first noun that becomes available. We assume this is so even though the postposition immediately following the noun may well be visible to readers at the point that the noun is fixated, indicating that another possible noun attachment host is likely to follow.[2] This immediate attachment strategy is the most natural given the assumption of strong incrementality that we have relied on in this thesis. It is also consistent with the data reported by Kamide and Mitchell (1996). However, we must point out here that further on-line data from a second experiment conducted by Yuki Kamide (personal communication) could be interpreted in terms of a *delay* strategy, in which immediate attachment does *not* take place. The second experiment differed from the first (Kamide and Mitchell, 1996) in that a different segmentation

[2]There are no spaces between words in Japanese script.

was used in the self-paced reading task. Below we summarize the segmentation regions used in the two experiments:[3]

Experiment 1: [RC] [N1] [P] [N2] ..... Verb

Experiment 2: [RC] [N1 P] [N2] ..... Verb

In both of Kamide and Mitchell's experiments, the relative clause was plausible as a modifier of only one of the two nouns. In the first experiment, a clear effect of plausibility was found in the reading time for N1, with faster reading times in cases where RC was a plausible modifier for N1 than when it was implausible. This indicates that immediate attachment of the relative clause was attempted. However, in the second experiment, where the postposition is visible in the same segment as N1, the difference was reduced so that it was no longer statistically significant, though the numerical trend was in the same direction as the experiment 1. This clearly shows that segmentation has an influence on the plausibility effect in this type of construction. However, it does not show definitively that the parser *delays* its attachment decision when the postposition is visible in the same region as N1 in normal reading, though it is certainly compatible with such an account. The results of Kamide and Mitchell's experiment 2 would also be compatible with an attachment strategy which is immediate, but in which the visible presence of the postposition weakens the commitment to the attachment, so that it no longer appears as a significant effect in the self-paced reading data. These issues could probably be resolved in an eye-tracking study, and the results which we will present in this chapter can be interpreted in terms of either a delaying or an immediate attaching parser. However, for present purposes, we will continue to assume that the attachment of the relative clause is immediate in the constructions under consideration. The introduction of delayed *attachment*, as opposed to the delayed *commitment* that we discussed in chapter 5, would require the introduction of some mechanism for buffering input, which is very much against the spirit of the strongly incremental model that we have been proposing in this thesis.

Dependency-preserving attachment of the postposition

The second assumption we would like to make is that in the type of construction exemplified in (8.2), the postposition is initially attached in a way which preserves the dependencies built

[3]The square brackets enclose experimental segments, "RC" stands for the relative clause, "N1" stands for the first noun, "N2" stands for the second noun, and "P" stands for the postposition.

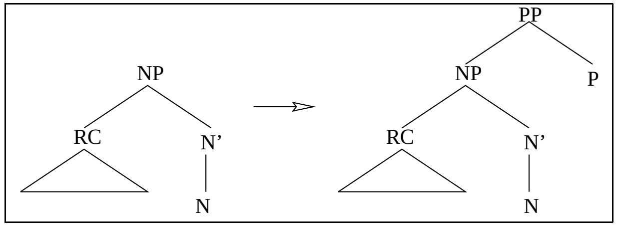


Figure 8.1: *Dependency preserving* attachment of the postposition in a complex NP. This corresponds to *left attachment* in terms of the model given in chapter 4.

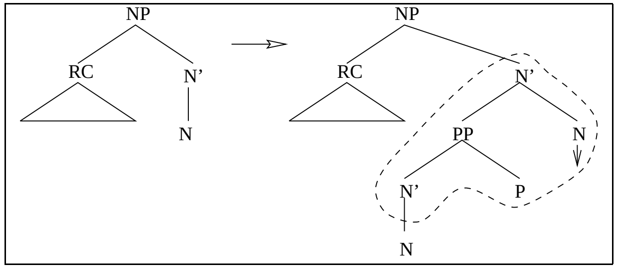


Figure 8.2: *Dependency-breaking* attachment of the postposition in a complex NP. This corresponds to an application of *tree-lowering* in the model given in chapter 4. The material inserted on the input of the postposition is enclosed in the dotted loop.

up until that point (cf. section 4.7, page 91). Given the initial immediate attachment of the relative clause to the first available noun, there are two possible ways in which the postposition can be attached; a *dependency preserving* attachment, corresponding to *left attachment* in the model developed in chapter 5, and a *dependency-breaking* attachment, corresponding to an application of the tree-lowering operation given in chapter 5. These two possibilities are shown in figures 8.1 and 8.2 respectively. As can be seen from figure 8.1, if the postposition is attached in accordance with the dependency-preserving strategy, the entire NP is taken as the left daughter of the resulting PP, and the dependency between the relative clause and the first noun is not broken. In contrast, the dependency-breaking attachment seen in

figure 8.2, involves the insertion of the PP at the intermediate N′[4], in such a way as the dependency between the relative clause and the first noun is broken. The off-line preference for high attachment shows that the dependency between the relative clause and the first noun is preferentially broken at some stage. The dependency-breaking attachment strategy outlined above would suggest that this dependency is broken almost immediately, as part of the attachment of the postposition. On the other hand, the dependency-preserving strategy allows that the dependency could broken at some other time, perhaps later in processing. The available on-line data suggest that the second alternative is the one that is taken by the human parser. This is because if the postposition is attached according to the dependency breaking strategy, a new dependency between the relative clause and the second noun will be formed immediately, and we would therefore expect to find early evidence for processing disruption in cases where the relative clause is not a plausible modifier for this second noun. However while Kamide and Mitchell (1996) did find such an effect, it was not local to the complex NP, but was rather found at the verb region at the end of the experimental sentences. This finding would be compatible with an account in which the reanalysis of the relative clause from the low to the high site does not occur as soon as the high site becomes available, but is delayed in some way.

8.1.2 Complex NPs and Thematic Structure

Given the assumptions that we have outlined above, that is, that the relative clause is initially attached to the first available (low) attachment site, and that the reanalysis to the high site occurs at some delay, it is reasonable to assume that the off-line preference for a high attachment of the relative clause reflects the results of reanalysis. Therefore, any moderating effect on this high attachment preference can be interpreted as a constraint on reanalysis.

The postposition *no* in (8.2) is similar to the English preposition *of*. Therefore, we may assume that, like *of*, it "transmits" a thematic role from the higher noun to the lower NP. Given the theory of thematic structure that we developed in chapter 5, then, reanalysing the relative clause from the low to the high site will be possible within the constraints of thematic monotonicity, since the relative clause will not lose any thematic dominators. On the other hand, if the postposition is substituted for another, that *assigns* rather than *transmits* a

[4] Recall that we have been assuming a maximum bar level of 1 for Japanese, so N′ is equivalent to NP.

thematic role to its complement, then this revision will violate thematic monotonicity. This is illustrated in Figure 8.3, which shows the thematic structure of the two complex NPs in (8.3):

(8.3) a. genjuumin ga odosita tankentai no taichou
natives NOM threatened expedition-force of commander
"The commander of the expedition force that the natives threatened"

b. genjuumin ga odosita tankentai kara-no taichou
natives NOM threatened expedition-force from commander
"The commander from the expedition force that the natives threatened"

We assume that in (8.3a), the higher noun *taicho* ("commander") assigns a thematic role to the lower NP headed by *tankentai* ("expedition force"). In (8.3b), however, we assume that the semantically contentful postposition [5] *kara-no* assigns a thematic role to the lower NP in its own right. Henceforth, we will call postpositions of the former type *non-thematic* postpositions, and those of the latter type *thematic* postpositions.

This has implications for the eventual interpretation of Japanese complex NPs such as (8.3). Subjects should retain their initial low-attachment more often when the complex NP contains a thematic role assigning postposition (as in 8.3b) than when it contains a postposition which does not assign a thematic role (as in 8.3a). This is because a thematic role assigning postposition delimits a thematic domain (see figure 8.3). In terms of the *thematic monotonicity* model developed in chapter 5, once the relative clause has been attached in the low site, the maximal projection of the thematic role assigning postposition will be a thematic dominator for the relative clause. Hence reanalysing the initially low-attached relative clause to the high site will be predicted to be dispreferred in such cases, since it entails that a thematic dominator is *lost* in cases such as (8.3b).

8.2 Norming Study

A norming study was conducted, allowing us to control for plausibility differences between the two readings of the experimental materials. We wrote a total of 56 materials similar to (8.3).

[5] Strictly speaking, *kara-no* might be more accurately called a a "postposition complex", because it is composed of two postpositions, *kara* and *no*. On its own, *kara* can only behave adverbially, and *no* is necessary to allow it to modify a noun.

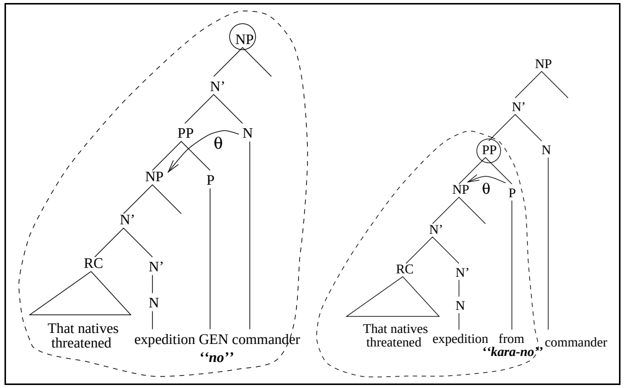


Figure 8.3: Syntactic structure of 8.3a and 8.3b. The circles indicate thematic assigner nodes, and the dotted loop indicates the domain within which the relative clause can be reanalysed within the constraints of thematic monotonicity

From each of these materials, we created two sentences, corresponding to the two readings of the ambiguity. For example, the two sentences for (8.3) were (Japanese translations of) *the natives threatened the expedition force* and *the natives threatened the commander*. Fifteen native speakers of Japanese from the University of Surugadai rated the plausibility of these sentences on a scale from 1 to 6. The sentences were presented in random order for each subject. The materials for both experiment 1 and experiment 2 were chosen from the set of sentences tested in the norming study, but unfortunately, the norming results were not available at the time when experiment 1 was run, so only experiment 2 was controlled for plausibility. In fact, the 24 materials chosen for experiment 1 turned out to be slightly biased in favour of the low attachment reading, which received a mean plausibility rating of 5.24, compared with 5.09 for the high attachment reading. This difference was analysed using a t-test, and was not significant by on the items analysis, but was significant on the subjects analysis; $t_1(14) = 3.08$, $p < .01$; $t_2(23) = 1.16$,ns. The materials for experiment 2 were selected so as to exhibit no significant difference between the plausibility ratings of the two readings.

8.3 Experiment 1

8.3.1 Method

Subjects

The subjects were twenty four native speakers of Japanese attending a course at the Park Language School in Sheffield.

Materials and Design

Materials were 24 pairs of complex NPs involving a relative clause and two NPs, similar to the example given in (8.3). In each pair, the two possible attachment sites for the relative clause were separated by either a non-thematic preposition (8.3a) or a thematic preposition (8.3b), and the two members of the pair differed only in the postposition separating the two noun attachment sites. To control for possible effects of animacy, we ensured that half of the 24 items included an animate noun in the high site and an inanimate noun in the low site, and half involved the reverse configuration. Henceforth, we will call the former configuration "animate-high" and the latter "animate-low". We used 26 fillers, which consisted of NPs with various types of internal structure, and various types of ambiguity. The materials are listed in appendix B, section B.1.

Procedure

The subjects saw individually randomized lists of materials, which were presented in printed booklets. The materials were rotated in a Latin Square design, so that each subject saw only one condition of each item. The fillers were interspersed among the materials in a random order, and in such a way that no two experimental items appeared adjacent to each other. Underneath each experimental or filler item, two sentences were included, which indicated the two possible interpretations of the item. For example, in the item given in (8.3), the two possible interpretations were indicated by (Japanese translations of) the following:[6]

a. The natives threatened the commander.

b. The natives threatened the expedition force.

[6] The a. and b. sentences were identical to the sentences used in the norming study.

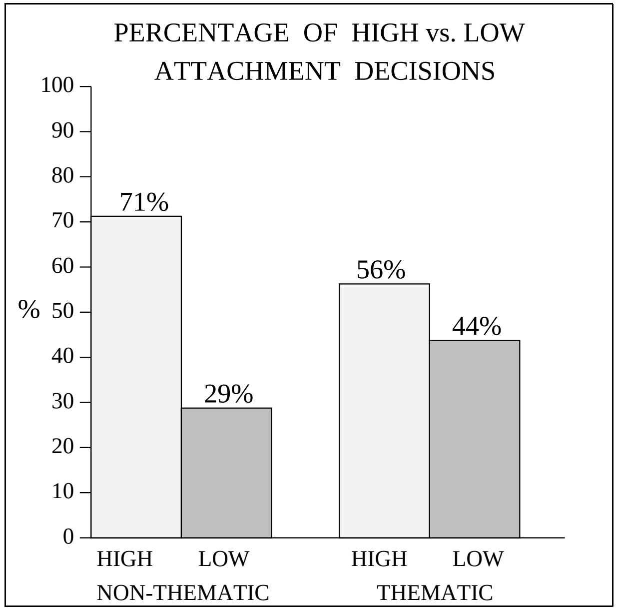


Figure 8.4:

Subjects had to indicate whether a. or b. corresponded to their first interpretation. The sentences were balanced so that the a. sentence indicated the high attachment interpretation half the time and vice versa.

8.3.2 Results

Overall, subjects chose the high attachment interpretation more often than the low attachment interpretation (63% high vs. 37% low). Comparing the proportion of high attachment decisions with 50% (i.e. the population mean expected on the null hypothesis) revealed that this preference was significant; $t_1(23) = 3.67$, $p < .01$; $t_2(23) = 3.93$, $p < .001$. However, there were significantly more high-attachment decisions in the non-thematic condition than in the thematic condition (71% vs. 56%); $t_1(23) = 3.03$, $p < .01$; $t_2(23) = 3.76$, $p < .01$. The proportion of high attachment decisions for the non-thematic condition was significantly greater than 50%; $t_1(23) = 4.96$, $p < .001$; $t_2(23) = 6.66$, $p < .001$. In contrast, the proportion of high attachment decisions for the thematic condition was not significantly greater than 50%; $t_1(23) = 1.24$, $p > .1$; $t_2(23) = 1.24$, $p > .1$. This indicates that the overall high attachment preference is driven by the non-thematic condition.

We analysed the effect of animacy, although this factor was not the focus of the experimental investigation, and was manipulated purely to create optimally balanced materials. *t*-tests comparing the animate-high condition with the animate-low condition revealed a greater proportion of high attachment decisions in the animate-low condition than the animate high condition (68% vs. 59%). However, this effect was significant by subjects only; $t_1(23) = 2.1$, $p < .05$, $t_2(11) = 1.22$, $p > .1$. An analysis of variance on the items data, taking thematicity as a within-items factor and animacy as a between-items factor revealed no significant interaction between the two factors; $F_2(1, 22) = 2.86$, $p > .1$.

8.3.3 Discussion

The study replicated the high-attachment preference found by Kamide and Mitchell (1996), but also demonstrated that postposition type has a reliable effect on attachment decisions in Japanese. Our findings are explicable if we assume that reanalyses which violate thematic monotonicity are dispreferred in relation to those which do not. This is also compatible with other theories which associate a processing cost with reanalysis outside the current domain

(De Vincenzi and Job (1993), Gilboy et al. (1995), Frazier and Clifton (1996)), although, as we argued in chapter 5, section 5.7, Frazier and Clifton's (1996) notion of thematic processing domain does not generalize adequately to left branching structures (see page 129 of this thesis).

However, the fact that no low-attachment preference was observed for the thematic condition indicates that the processor does sometimes reanalyse across a thematic domain boundary, even in the absence of a syntactic or semantic cue (note especially that the plausibility bias would, if anything, be expected to favour the *low* attachment). This indicates that although reanalysis which violates thematic monotonicity is dispreferred in relation to reanalysis which does not, it is nevertheless not true to say that the processor avoids preforming such non-monotonic reanalyses at all costs. We will discuss this issue further in the general discussion section.

8.4 Experiment 2

Experiment 1 left us with further questions, which were addressed in a second experiment. Experiment 1 confirmed that the reanalysis of relative clauses is sensitive to domains that can be defined in terms of *thematic assigner* nodes. However, a further question is that of whether domains differ in their relative "strengths", depending on their syntactic characteristics. In particular, we sought to test whether the presence or absence of a *clause boundary* between the two possible attachment sites of the relative clause has an effect on the final interpretation. Experiment 1 has demonstrated that, assuming an initial low attachment preference for the relative clause, a postpositional thematic domain boundary has a "containing effect" on the relative clause, keeping it in the lower attachment site more often than occurs when such a boundary is absent. However, a plausible hypothesis would be that this containing effect is "stronger" in cases where the thematic domain boundary is clausal than in cases where it is merely postpositional. This would be predicted by Gibson et al's (1996) *Predicate Proximity* Hypothesis, for example, and would be a natural assumption to make linguistically, since the clausal, but not the postpositional domain can be assumed to be a "Complete Functional Complex" (Chomsky, 1986b), a unit which is known to act as a domain for the purposes of binding and other syntactic phenomena. Such a finding would have a consequence on our model, in that it would require us to define an extra level of structure, for example, differentiating between clausal and non-clausal thematic nodes.

8.4.1 Method

Subjects

Subjects were twenty four native speakers of Japanese who were attending a course at the Institute of Applied Language Studies at the University of Edinburgh.

Materials and Design

The materials consisted of 24 triples of complex NPs. Each triple consisted of three conditions, the first two of which were a *non-thematic* condition and a *thematic condition*, which together correspond to the two conditions of experiment 1. The third condition was a *clausal* condition, in which the two possible attachment sites are separated by a clause boundary, created by modifying the higher noun with a relative clause. This third condition is illustrated in example (8.4c) below, and a possible tree structure for its low attachment option is given in figure 8.5. The full list of materials appears in appendix B, section B.2.

(8.4) a. genjuumin ga odosita tankentai no taichou
natives NOM threatened expedition force of commander
"The commander of the expedition force that the natives threatened"

b. genjuumin ga odosita tankentai kara-no taichou
natives NOM threatened expedition from commander
"The commander from the expedition force that the natives threatened"

c. genjuumin ga odosita tankentai wo hikiita taichou
natives NOM threatened expedition ACC led commander
"The commander who led the expedition force that the natives threatened"

Note that the clausal condition involves two relative clauses; the initial relative clause, whose attachment is ambiguous, and a second relative clause, which modifies the higher of the two attachment sites. As in experiment 1, 12 of the 24 items were animate-high condition, and 12 were animate-low. In all the animate-high items in the clausal condition, this second relative clause was a *subject relative*; that is, it included a subject gap, coindexed with the (animate) higher noun. In all the animate-low items, the second relative clause was a *non-subject* relative; that is, it included a non-subject gap, coindexed with the (inanimate) higher

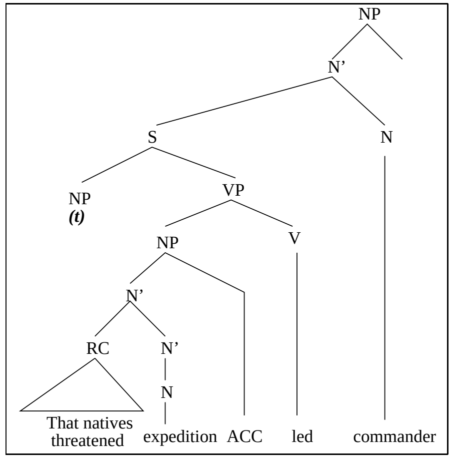


Figure 8.5:

noun. This was necessary because of the strong preference for transitive verbs to have animate subjects in Japanese.

The materials were chosen to reduce to a minimum the mean plausibility difference between the high and low attachment interpretations for each item, while still maintaining the equal balance of animate-high and animate-low items. The mean plausibility ratings for the high and low interpretations for experiment 2 were 5.35 and 5.37 respectively. This difference was non-significant by both subjects and items (both t's < 1). This resulted in a set of materials which shared 13 out of 24 items in common with those of experiment 1, and whose overall plausibility was almost exactly balanced between the high and the low interpretations. The experiment used the same fillers as experiment 1.

Procedure

The procedure was identical to that of experiment 1.

8.4.2 Results

The results are summarized in figure 8.6.

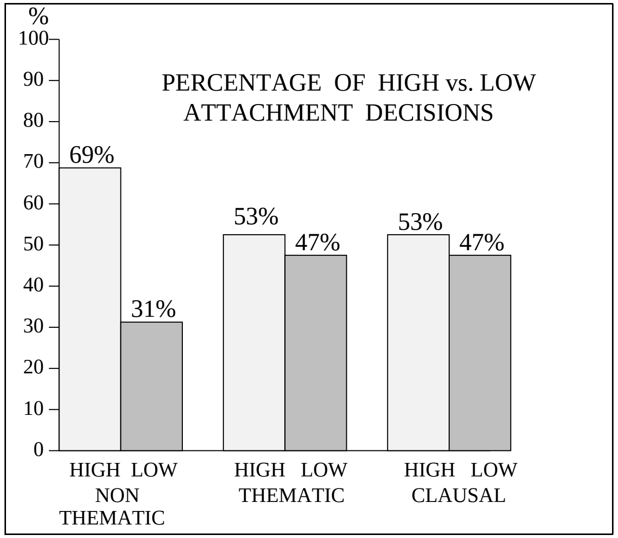


Figure 8.6:

Overall there was a preference for high attachment (58% high vs. 42% low). t-tests comparing the overall proportion of high attachments with 50% revealed that this preference was significant by subjects but not by items; $t_1(23) = 2.5$, $p < .05$; $t_2(23) = 1.47$, $p = 0.16$. Planned contrasts revealed that there were significantly more high-attachment decisions in the non-thematic condition than in the thematic condition (69% vs. 53%); $t_1(23) = 4.53$, $p < .001$; $t_2(23) = 3.82$, $p < .001$, and also that there were significantly more high attachment decisions in the non-thematic condition than in the clausal condition (69% vs. 53%); $t_1(23) = 4.17$, $p < .001$; $t_2(23) = 2.81$, $p < .02$. However, there was no difference between the thematic and the clausal conditions—in fact the means were numerically identical (53% vs. 53%); $t_1 = t_2 = 0$, $p = 1$.

As in Experiment 1, although animacy was manipulated as a control, and was not the intended focus of the study, we analysed the effect of this factor. Again, more high attachment decisions were taken in the animate-low condition than in the animate-high condition (80% vs. 36%), but this time, the effect was strongly significant by both subjects and items; $t_1(23) = 8.04$, $p < .001$; $t_2(22) = 7.23$, $p < .001$. To test whether animacy interacted with the effect of thematicity, we conducted a mixed 2×3 analysis of variance on the items data, with thematicity as a within-items factor, and animacy as a between-items factor. As expected, this yielded significant main effects of animacy and thematicity, but no interaction between the two factors ($F_2(2, 44) < 1$). Animacy also affected the attachment bias. Separate t-tests were calculated for the animate-high and animate-low conditions, comparing the number of high attachments against 50%. This revealed an overall high attachment preference for the animate-low items; $t_1(23) = 6.70$, $p < .001$; $t_2(11) = 9.34$, $p < .001$, and an overall low attachment preference for the animate-high items $t_1(23) = 4.45$, $p < .001$; $t_2(11) = 2.70$, $p < .05$. Figure 8.7 provides a summary of the effects of animacy and thematicity.

8.4.3 Discussion

Experiment 2 replicated the finding of Experiment 1 that a thematic postposition neutralizes the high attachment preference found with the non-thematic postposition. However, Experiment 2 also shows that a clausal thematic domain has precisely the same effect as a postpositional thematic domain. This means that we do not have evidence for two different types of thematic domains. However, an unexpected result of Experiment 2 is the very strong

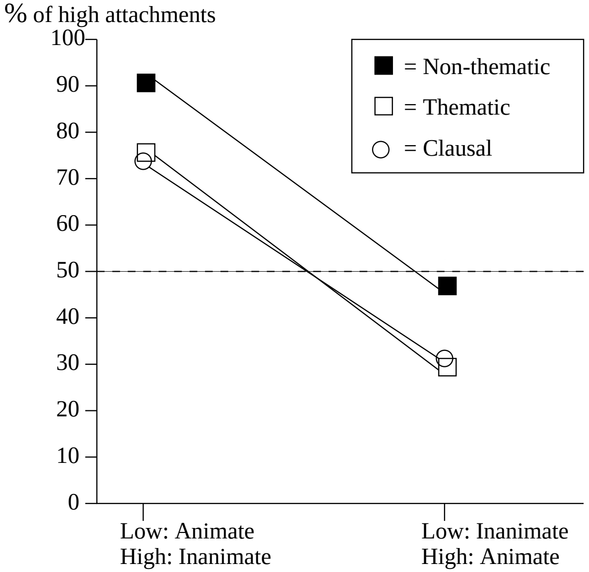


Figure 8.7: Experiment 2: Effects of animacy and thematicity

effect of animacy, which was not found to anything like the same extent in Experiment 1. We will discuss some reasons for this difference in section 8.5.3. A possible reason for the effect of animacy is that animate nouns are more "conceptually salient" than inanimate nouns, thus making them less in need of the referential support provided by modification. The effect might also be the result of the particular choice of materials that we made in the experiment, where many of the animate nouns can be interpreted as picking out a unique referent (e.g. "teacher", "company president"), and thus may resist modification for presuppositional reasons (Crain and Steedman, 1985; Altmann and Steedman, 1988).[7] This might be the explanation for a similar effect which has been found in German (Hemforth, Konieczny, and Scheepers, 1994b), where the authors found a high attachment preference for relative clauses in complex NPs similar to those discussed in this chapter. In this German study, a marked preference was only found in cases where the higher noun was non-human and the lower noun was human. Clearly, further research is required to locate the source of the animacy effect, and to determine *when* in the comprehension process animacy has its effect, a question which cannot be answered using the questionnaire technique.

8.5 General Discussion

8.5.1 Thematicity

These two experiments have shown that altering thematic characteristics of the complex NP has an effect on the attachment preference of the relative clause. However, it is not the case that the presence of a thematic domain boundary, whether clausal or postpositional, *blocks* the high attachment of the relative clause. Indeed, the data from experiment 2 show that, under favourable semantic conditions, there can be an overall high attachment preference for the relative clause even where a clausal or postpositional thematic domain boundary intervenes between the two possible attachment sites. What the data of experiment 2 appears to show is that the processor prefers to maintain the initial *low* attachment of the relative clause only in cases where *both* thematic *and* the animacy constraints favour it.

An alternative explanation of the effect of postposition type found in experiment 1 was suggested to us by Janet Fodor (p.c.). This explanation depends on the observation that

[7] Chuck Clifton pointed this out to us.

the preference for recent, low attachment is stronger for shorter phrases than for longer ones (Frazier, 1978; Frazier and Fodor, 1978). Consider this claim in relation to the complex NPs that we examined in experiment 1.

(8.3) a. genjuumin ga odosita tankentai no taichou
natives NOM threatened expedition-force of commander
"The commander of the expedition force that the natives threatened"

b. genjuumin ga odosita tankentai kara-no taichou
natives NOM threatened expedition from commander
"The commander from the expedition force that the natives threatened"

Recal figures 8.1 and 8.2 on page 197. The thematic postposition *kara-no* is longer than the non-thematic postposition *no*. Therefore, *no* might be expected to exhibit a preference for the low (dependency-breaking) attachment in 8.2, which would result in a *high* attachment of the relative clause. Conversely, the longer *kara-no* might be expected to exhibit a preference for the high (dependency-preserving) attachment illustrated in figure 8.1, resulting in a *low* attachment of the relative clause.[8] In order to test this claim, it would be necessary to manipulate length and thematicity of the postposition independently. Unfortunately, this manipulation is impossible in Japanese, since any noun-modifying thematic postposition is guaranteed to be longer than the non-thematic postposition *no*. This is because, (as far as we are aware), thematic postpositions must include *no* as a suffix to allow them to modify nouns. In order to make a satisfactory test of this claim, we would therefore need to find a head-final, left branching language in which non-thematic postpositions can be longer than thematic postpositions.[9]

However, as we have mentioned above, if the postposition is attached in the *low* (or *dependency-breaking*) manner, then we would expect an immediate effect of plausibility at the point where the higher noun is read (i.e. longer reading times where the relative clause is an implausible modifier of the noun than when it is plausible). Kamide and Mitchell (1996)

[8]Recall the discussion of the trade-off between dependency preservation and recency in chapter 4, section 4.7, page 91.

[9]The results of experiment 2 allow us to reject one particular version of this length hypothesis, in which the strength of the high attachment preference of the postposition (and thus the likelihood of *low* attachment of the relative clause) correlates linearly with the length of the material separating the two possible noun hosts. In this case, we would expect to find a difference between the clausal condition (with an average of 4 characters intervening between the two nouns) and the thematic postposition condition (where the number of intervening characters is always 3). However, this difference was not found.

found such a plausibility effect in their on-line study, but it appeared at the final verb region, and not at the region of the higher noun.

8.5.2 High Attachment Bias

As for the *reason* for the overall high attachment preference, which was found in both experiment 1 and experiment 2, it is not clear how current theories can account for this. One possible explanation could be some version of *Relativised Relevance* (Frazier, 1990b), which predicts a preference for associating a modifier with the "main assertion of the sentence". We note that, as our study examined NPs, which are not sentences and do not denote "assertions"[10], this principle would have to apply either to non-overt clausal structure or to apply to the "main content" of the utterance, whatever its categorial expression or semantic type. Gibson et al. (1996) could account for the overall high preference, assuming that their *Predicate Proximity* preference is stronger than their *Recency* preference. However, their model could not account for the effect of postposition type found here, unless a secondary principle of predicate proximity were to be invoked that could deal with the predicates associated with pre-/postpositions.

8.5.3 Animacy and Plausibility

One striking aspect of the data is the difference in the animacy effect between the two experiments—in Experiment 2, the effect of animacy was very highly significant, and was strong enough to alter the attachment bias from high to low. In experiment 1, by contrast, animacy had no significant effect. There are several possible reasons for this difference; it could reflect a difference between the materials of experiments 1 and 2 (recall that the two experiments shared only twelve out of 24 items). Otherwise, it could be due to the different subjects who took part in the two experiments, or one or other of the experiments may have produced atypical results by chance. We believe that the the first explanation is correct; the difference is a result of using different materials in the two experiments. Evidence for this can be found if we separate the items which the two experiments share in common from those which were only used in one experiment. Firstly, a Pearson's correlation test showed that the total number of high attachment decisions for the items shared between experiment 1 and

[10]We take "assertion" to be synonymous with "proposition".

experiment 2 correlated significantly across the two experiments ($r = .71$, $p < .01$). This indicates that the two subject groups employed similar preferences in choosing their interpretations. If we analyse the data of experiment 1 considering only those items which were shared in common with experiment 2, then the pattern of results comes to resemble those of experiment 2; more high attachment decisions were taken in the animate-low items than in the animate-high items (75% vs. 52%), and this difference was significant by a t-test on the items analysis; $t_2(11) = 3.72$, $p < .01$. For the remaining items of experiment 1 (i.e. those which were not shared with experiment 2), the number of high attachment decisions in the animate-low items did not differ significantly from the number of high attachment decisions in the animate-high items; in fact, the numerical trend was in the opposite direction, with high attachment being chosen in 55% of the animate-low items, as against 64% for the animate high items ($t_2(10) < 1$).

This suggests that the source of the difference in animacy effects between the two experiments may lie partly in the items which they did not share in common. Recall that the materials for experiment 1 were not controlled for plausibility. The items which were removed from experiment 1 were those which were judged to exhibit the greatest plausibility difference between the high attachment and low attachment reading in the norming study, and these were replaced by items whose readings differed only minimally in plausibility between the two readings, to create the materials for experiment 2. The number of high attachment decisions for the non-thematic condition of these items (i.e. those which were included in experiment 1 but excluded from experiment 2) was found, by a Pearson's correlation test, to correlate with the plausibility of their respective high attachment readings, ($r = .63, p < .05$),[11]. By contrast, the corresponding correlation test for those items of experiment 1 which were *not* excluded from experiment 2 failed to reach significance ($r = .23$, n.s.). We made a number of other tests, looking for correlations between attachment decisions in both experiments and plausibility, but none of these yielded significant correlations.[12] This suggests that plausibility

[11]Interestingly, the number of high attachment decisions for the non-thematic condition did not correlate with the plausibility of *low* attachment readings ($r = .28$, ns.) for these items. Furthermore, the thematic condition did not correlate significantly with the plausibility of either the high ($r = .36$, ns) nor the low attachment ($r = -0.14$, ns) interpretations. The difference between the thematic and non-thematic conditions may be the result of a semantic effect similar to the Construal principle (Frazier and Clifton, 1996). However, we have no explanation for the fact that only the plausibility of the *high* reading is involved in the correlation.

[12]These tests considered data from each experiment as a whole, and tested for correlations between the number of high attachment decisions and a number of values calculated from the plausibility data: plausibility

only affected the subjects' interpretation of the materials where there was a large difference in plausibility between the two readings, and moreover, that it was only when such large differences did exist, animacy failed to have an effect.

8.6 Conclusions

From the two questionnaire studies reported in this chapter, we can conclude that the interpretation of a Japanese relative clause is affected by the syntactic and thematic details of the complex NP into which it is attached. Specifically, while there is an overall preference for the relative clause to be attached to the higher of the two possible attachment sites, the presence of a thematic domain boundary between the sites significantly decreases the number of high attachment decisions made. However, this effect is identical whether the thematic domain is clausal or postpositional, at least as far as the off-line preferences tested in this chapter are concerned. We have also seen that animacy has an important role to play. However, it is necessary to conduct further on-line studies in order to understand the way in which these factors manifest themselves in the language comprehension process.

of the high attachment reading, plausibility of the low attachment reading, difference in plausibility between the two readings, and mean plausibility.

Chapter 9

Conclusions

In chapter 2, we pointed out that most existing research in human syntactic processing has been concerned with the issue of how people initially resolve ambiguities, rather than with the complementary question of how people recover when the initially chosen resolution of the ambiguity turns out to be wrong. The main contribution of this thesis has been, we believe, to demonstrate that reanalysis is indeed a viable subject of research, and that it is amenable to investigation through theoretical, computational and experimental methods.

On the theoretical side, we have discussed in some detail the components of the reanalysis process, and shown how models may vary with respect to which aspects of the process they constrain. Concentrating on the notion of representation preservation, we have shown how the insights of D-theory (Marcus, Hindle, and Fleck, 1983) can be used as a basis of a generalized monotonicity constraint, which can be applied to various types of representation, including *constituent structure* (chapter 4), and thematic structure (chapter 5).

Computationally, we have shown the importance of considering not only *which* structures the parser can easily revise, but also *how* the parser revises them. Such issues have to be covered in theories of reanalysis, because they are directly related to the notion of reanalysis ambiguity, as discussed in chapter 4. Reanalysis ambiguity must be a crucial issue, since, considering the degree of ambiguity present in natural language, it is highly likely that, at the point where reanalysis is attempted, there will be more than one previous choice point which could conceivably be reanalysed in a number of different ways. In fact in purely head final languages, and given certain assumptions concerning incremental processing, *all* instances of structural ambiguity in such languages can be argued to be reanalysis ambiguity. A further

contribution on the computational side has been the development of a general method for incremental processing, and a method of constraining reanalysis search based on *indexing* (chapter 6).

In terms of experimental research, our main contribution has been to consider the experimental investigation of *reanalysis ambiguity*. As far as we are aware, the experiments reported in chapter 7 are the first to consider this issue in English, although we know of experimental research on reanalysis ambiguity in Japanese (Hirose, 1997). While we have not resolved the issue of a good *on-line* measure of preferences in reanalysis ambiguity, the results of the *off-line* measures presented in that chapter indicate that the matter requires further investigation. Our other experimental contribution has been to establish the existence of construal-type effects on relative clause attachment in Japanese (chapter 8). Although we have argued that the effect reported in that chapter should be seen as an effect of reanalysis, we require on-line investigations to establish the time course of the effect.

9.1 Future Directions

Below we discuss a number of ways in which the research reported in this thesis could be extended in the future.

9.1.1 Experimental Investigation of Intuitive Effects

A weakness of many of the theories of reanalysis which have been proposed is that they rely to a great extent on *intuitive* data that have not been experimentally confirmed. This is particularly true in the case of claims made by, for example, Pritchett (1988; 1992), in which easy and hard garden paths are differentiated in terms of whether or not the reanalysis involved is available for conscious introspection—the question of whether or not a garden path reaches consciousness is not something that is likely to be testable by standard experimental techniques. However, reanalysis theories, including that proposed in this thesis, can be seen as predicting that, all other things being equal, certain types of reanalysis cause more processing difficulty than others. This question is independent of whether or not the reanalysis in question causes *conscious* processing difficulty, and should be amenable to study using established experimental techniques. This would involve comparing minimal pairs of sentences such as the following, from chapter 2:

(9.1) While the wedding guests ate the cake was still being decorated.

(9.2) The wedding guests saw the cake was still being decorated.

Another related question concerns the extent to which the difficulty of garden paths can be divided into two discrete categories. Many theories of reanalysis have tacitly made this assumption in their distinctions between, for example, conscious and unconscious reanalysis, or, in terms of the present thesis, the distinction between monotonic and non-monotonic reanalysis. However, even if these theoretical distinctions do have a systematic effect on reanalysis difficulty, there remains the question of how they relate to other factors which are known to affect reanalysis. For example, can reanalysis difficulty be broadly divided into two categories describable in structural terms, such as the monotonic/non-monotonic distinction, or is reanalysis difficulty more accurately described as a *continuum* driven by graded phenomena such as, for example, the difference in statistical preference between the two readings before and after reanalysis, within which structural factors may play a part. These are interesting and important questions, which can only be addressed using experimental methods.

9.1.2 Further Extensions of Computational Model

Although the use of computational modelling has been useful in identifying basic mechanisms for incremental processing and reanalysis, our work in this area has been limited by our use of a rather small scale grammar. Extending the system to run on a wide coverage grammar would vastly increase the degree of ambiguity with which the parser would have to cope, and this would almost certainly highlight unforeseen areas in which the grammar interacts with the processing model to produce surprising results. This would lead to opportunities for further development.

Appendix A

Experimental Materials from chapter 7

A.1 Self-Paced Reading Experiment

The following is a list of materials for the self-paced reading study reported as "Experiment 1" in chapter 7. The conditions are to be read as follows: "AL" is "Ambiguous Low", "AH" is "Ambiguous High", "UL" is "Unambiguous Low" and "UH" is "Unambiguous High". The symbol "#" marks the region boundary, and "Q" marks the comprehension question.

1. AL: The investigation revealed the councillor who accepted the spending cuts # had not justified themselves.
   Q: Were the spending cuts justified?
   AH: The investigation revealed the councillor who accepted the spending cuts # had not justified himself.
   Q: Was the councillor justified?
   UL: The investigation revealed the councillor who accepted that the spending cuts # had not justified themselves.
   Q: Were the spending cuts justified?
   UH: The investigation revealed that the councillor who accepted the spending cuts # had not justified himself.
   Q: Was the councillor justified?
2. AL: The children's home accepted the girl who revealed her tormenters # had been injecting themselves.
   Q: Had the girl been taking drugs?
   AH: The children's home accepted the girl who revealed her tormenters # had been injecting herself.
   Q: Had the tormenters been taking drugs?

UL: The children's home accepted the girl who revealed that her tormenters # had been injecting themselves.

Q: Had the girl been taking drugs?

UH: The children's home accepted that the girl who revealed her tormenters # had been injecting herself.

Q: Had the tormenters been taking drugs?

3. AL: The editor accepted the story which revealed the corrupt politicians # had contradicted each other.

Q: Was the story self-contradictory?

AH: The editor accepted the story which revealed the corrupt politicians # had contradicted itself.

Q: Had the politicians been contradicting each other?

UL: The editor accepted the story which revealed that the corrupt politicians # had contradicted each other.

Q: Was the story self-contradictory?

UH: The editor accepted that the story which revealed the corrupt politicians # had contradicted itself.

Q: Had the politicians been contradicting each other?

4. AL: The detective forgot the witnesses who recognised the murderer # had disguised himself.

Q: Were the witnesses in disguise?

AH: The detective forgot the witnesses who recognised the murderer # had disguised themselves.

Q: Was the murderer in disguise?

UL: The detective forgot the witnesses who recognised that the murderer # had disguised himself.

Q: Were the witnesses in disguise?

UH: The detective forgot that the witnesses who recognised the murderer # had disguised themselves.

Q: Was the murderer in disguise?

5. AL: The general recalled the troops who found the enemy spy # had shot himself.

Q: Had the troops committed suicide?

AH: The general recalled the troops who found the enemy spy # had shot themselves.

Q: Had the enemy spy committed suicide?

UL: The general recalled the troops who found that the enemy spy # had shot himself.

Q: Had the troops committed suicide?

UH: The general recalled that the troops who found the enemy spy # had shot themselves.

Q: Had the enemy spy committed suicide?

6. AL: The pensioners recalled the wartime singer who found the German pilots # had poisoned themselves.

Q: Had the singer taken poison?

AH: The pensioners recalled the wartime singer who found the German pilots # had poisoned herself.

Q: Had the pilots taken poison?

UL: The pensioners recalled the wartime singer who found that the German pilots # had poisoned themselves.

Q: Had the singer taken poison?

UH: The pensioners recalled that the wartime singer who found the German pilots # had poisoned herself.

Q: Had the pilots taken poison?

7. AL: The private detective found the informants who recalled the suspect # had disguised himself.

Q: Had the suspect been in disguise?

AH: The private detective found the informants who recalled the suspect # had disguised themselves.

Q: Had the informants been in disguise?

UL: The private detective found the informants who recalled that the suspect # had disguised himself.

Q: Had the suspect been in disguise?

UH: The private detective found that the informants who recalled the suspect # had disguised themselves.

Q: Had the informants been in disguise?

8. AL: The professor acknowledged the students who understood the new theory # had contradicted itself.

Q: Was the new theory self-contradictory?

AH: The professor acknowledged the students who understood the new theory # had contradicted each other.

Q: Had the students been inconsistent with each other?

UL: The professor acknowledged the students who understood that the new theory # had contradicted itself.

Q: Was the new theory self-contradictory?

UH: The professor acknowledged that the students who understood the new theory # had contradicted each other.

Q: Had the students been inconsistent with each other?

9. AL: The social worker understood the teenagers who acknowledged their gang leader # couldn't express himself.

Q: Was the gang leader bad at expressing himself?

AH: The social worker understood the teenagers who acknowledged their gang leader # couldn't express themselves.

Q: Were the teenagers bad at expressing themselves?

UL: The social worker understood the teenagers who acknowledged that their gang leader # couldn't express himself.

Q: Was the gang leader bad at expressing himself?

UH: The social worker understood that the teenagers who acknowledged their gang leader # couldn't express themselves.

Q: Were the teenagers bad at expressing themselves?

10. AL: The editor found the photographers who saw the cabinet minister # had disgraced himself.

Q: Had the cabinet minister done something disgraceful?

AH: The editor found the photographers who saw the cabinet minister # had disgraced themselves.

Q: Had the photographers done something disgraceful?

UL: The editor found the photographers who saw that the cabinet minister # had disgraced himself.

Q: Had the cabinet minister done something disgraceful?

UH: The editor found that the photographers who saw the cabinet minister # had disgraced themselves.

Q: Had the photographers done something disgraceful?

11. AL: The psychiatrist accepted the male patient who forgot his wife # couldn't look after herself.

Q: Was the man's wife bad at looking after herself?

AH: The psychiatrist accepted the male patient who forgot his wife # couldn't look after himself.

Q: Was the patient bad at looking after himself?

UL: The psychiatrist accepted the male patient who forgot that his wife # couldn't look after herself.

Q: Was the man's wife bad at looking after herself?

UH: The psychiatrist accepted that the male patient who forgot his wife # couldn't look after himself.

Q: Was the patient bad at looking after himself?

12. AL: The inspector recognised the builders who accepted the young apprentice # would be a danger to himself.

Q: Was the young apprentice a likely to harm himself?

AH: The inspector recognised the builders who accepted the young apprentice # would be a danger to themselves.

Q: Were the builders likely to harm themselves?

UL: The inspector recognised the builders who accepted that the young apprentice # would be a danger to himself.

Q: Was the young apprentice likely to harm himself?

UH: The inspector recognised that the builders who accepted the young apprentice # would be a danger to themselves.

Q: Were the builders likely to harm themselves?

13. AL: The judge accepted the witness who recognised the suspects # had been deceiving themselves.

Q: Had the suspects been honest with themselves?

AH: The judge accepted the witness who recognised the suspects # had been deceiving himself.

Q: Had the witness been honest with himself?

UL: The judge accepted the witness who recognised that the suspects # had been deceiving themselves.

Q: Had the suspects been honest with themselves?

UH: The judge accepted that the witness who recognised the suspects # had been deceiving himself.

Q: Had the witness been honest with himself?

14. AL: The tabloids revealed the informers who recalled the old politician # had exposed himself.

Q: Had the informers made an exhibition of themselves?

AH: The tabloids revealed the informers who recalled the old politician # had exposed themselves.

Q: Had the politician made an exhibition of himself?

UL: The tabloids revealed the informers who recalled that the old politician # had exposed himself.

Q: Had the informers made an exhibition of themselves?

UH: The tabloids revealed that the informers who recalled the old politician # had exposed themselves.

Q: Had the politician made an exhibtion of himself?

15. AL: The shop steward understood the workers who accepted the new manager # could not commit himself.

Q: Was the new manager unable to make a commitment?

AH: The shop steward understood the workers who accepted the new manager # could not commit themselves.

Q: Were the workers unable to make a commitment?

UL: The shop steward understood the workers who accepted that the new manager # could not commit himself.

Q: Was the new manager unable to make a commitment?

UH: The shop steward understood that the workers who accepted the new manager # could not commit themselves.

Q: Were the workers unable to make a commitment?

16. AL: The director understood the administrator who accepted the incompetent trainees # couldn't explain themselves.

Q: Were the trainees bad communicators?

AH: The director understood the administrator who accepted the incompetent trainees # couldn't explain himself.

Q: Was the administrator a bad communicator?

UL: The director understood the administrator who accepted that the incompetent trainees # couldn't explain themselves.

Q: Were the trainees bad communicators?

UH: The director understood that the administrator who accepted the incompetent trainees # couldn't explain himself.

Q: Was the administrator a bad communicator?

17. AL: The examiner accepted the student who understood the Latin authors # had repeated themselves.

Q: No question.

AH: The examiner accepted the student who understood the Latin authors # had repeated himself.

Q: No question.

UL: The examiner accepted the student who understood that the Latin authors # had repeated themselves.

Q: No question.

UH: The examiner accepted that the student who understood the Latin authors # had repeated himself.

Q: No question.

18. AL: The personnel officer accepted the applicant who understood the policies # had justified themselves.

Q: No question.

AH: The personnel officer accepted the applicant who understood the policies # had justified himself.

Q: No question.

UL: The personnel officer accepted the applicant who understood that the policies # had justified themselves.

Q: No question.

UH: The personnel officer accepted that the applicant who understood the policies # had justified himself.

Q: No question.

19. AL: The psychoanalyst saw the cult members who recalled the possessed boy # couldn't control himself.

Q: No question.

AH: The psychoanalyst saw the cult members who recalled the possessed boy # couldn't control themselves.

Q: No question.

UL: The psychoanalyst saw the cult members who recalled that the possessed boy # couldn't control himself.

Q: No question.

UH: The psychoanalyst saw that the cult members who recalled the possessed boy # couldn't control themselves.

Q: No question.

20. AL: The camp organiser recalled the volunteer who saw the teenagers # had enjoyed themselves.

Q: No question.

AH: The camp organiser recalled the volunteer who saw the teenagers # had enjoyed himself.

Q: No question.

UL: The camp organiser recalled the volunteer who saw that the teenagers # had enjoyed themselves.

Q: No question.

UH: The camp organiser recalled that the volunteer who saw the teenagers # had enjoyed himself.

Q: No question.

21. AL: The CIA revealed the farmers who saw the alien # had surrendered itself.

Q: No question.

AH: The CIA revealed the farmers who saw the alien # had surrendered themselves.

Q: No question.

UL: The CIA revealed the farmers who saw that the alien # had surrendered itself.

Q: No question.

UH: The CIA revealed that the farmers who saw the alien # had surrendered themselves.

Q: No question.

22. AL: The audience saw the band members who acknowledged their promoter # had intoxicated himself.

Q: No question.

AH: The audience saw the band members who acknowledged their promoter # had intoxicated themselves.

Q: No question.

UL: The audience saw the band members who acknowledged that their promoter # had intoxicated himself.

Q: No question.

UH: The audience saw that the band members who acknowledged their promoter # had intoxicated themselves.

Q: No question.

23. AL: The publisher recognised the author who acknowledged his secretaries # had disgraced themselves.

Q: No question.

AH: The publisher recognised the author who acknowledged his secretaries # had disgraced himself.

Q: No question.

UL: The publisher recognised the author who acknowledged that his secretaries # had disgraced themselves.

Q: No question.

UH: The publisher recognised that the author who acknowledged his secretaries # had disgraced himself.

Q: No question.

24. AL: The bishop acknowledged the priest who accepted the sinners # had reformed themselves.

Q: No question.

AH: The bishop acknowledged the priest who accepted the sinners # had reformed himself.

Q: No question.

UL: The bishop acknowledged the priest who accepted that the sinners # had reformed themselves.

Q: No question.

UH: The bishop acknowledged that the priest who accepted the sinners # had reformed himself.

Q: No question.

25. AL: The company boss acknowledged the department which accepted the young graduates # had proved themselves.

Q: No question.

AH: The company boss acknowledged the department which accepted the young graduates # had proved itself.

Q: No question.

UL: The company boss acknowledged the department which accepted that the young graduates # had proved themselves.

Q: No question.

UH: The company boss acknowledged that the department which accepted the young graduates # had proved itself.

Q: No question.

26. AL: The butler saw the prince who revealed his mistress # had hanged herself.

Q: No question.

AH: The butler saw the prince who revealed his mistress # had hanged himself.

Q: No question.

UL: The butler saw the prince who revealed that his mistress # had hanged herself.

Q: No question.

UH: The butler saw that the prince who revealed his mistress # had hanged himself.

Q: No question.

27. AL: The war veterans recalled the officer who revealed the traitors # had drowned themselves.

Q: No question.

AH: The war veterans recalled the officer who revealed the traitors # had drowned himself.

Q: No question.

UL: The war veterans recalled the officer who revealed that the traitors # had drowned themselves.

Q: No question.

UH: The war veterans recalled that the officer who revealed the traitors # had drowned himself.

Q: No question.

28. AL: The explorers recalled the tribesmen who revealed the witch-doctor # had sacrificed himself.

Q: No question.

AH: The explorers recalled the tribesmen who revealed the witch-doctor # had sacrificed themselves.

Q: No question.

UL: The explorers recalled the tribesmen who revealed that the witch-doctor # had sacrificed himself.

Q: No question.

UH: The explorers recalled that the tribesmen who revealed the witch-doctor # had sacrificed themselves.

Q: No question.

29. AL: The scientist found the Eskimos who recalled the famous explorer # had no confidence in himself.

Q: No question.

AH: The scientist found the Eskimos who recalled the famous explorer # had no confidence in themselves.

Q: No question.

UL: The scientist found the Eskimos who recalled that the famous explorer # had no confidence in himself.

Q: No question.

UH: The scientist found that the Eskimos who recalled the famous explorer # had no confidence in themselves.

Q: No question.

30. AL: The detective found the children who recalled the kidnapper # despised himself.

Q: No question.

AH: The detective found the children who recalled the kidnapper # despised themselves.

Q: No question.

UL: The detective found the children who recalled that the kidnapper # despised himself.

Q: No question.

UH: The detective found that the children who recalled the kidnapper # despised themselves.

Q: No question.

31. AL: The newspaper report revealed the boy who found the escaped prisoners # had injured themselves.

Q: No question.

AH: The newspaper report revealed the boy who found the escaped prisoners # had injured himself.

Q: No question.

UL: The newspaper report revealed the boy who found that the escaped prisoners # had injured themselves.

Q: No question.

UH: The newspaper report revealed that the boy who found the escaped prisoners # had injured himself.

Q: No question.

32. AL: The article revealed the policeman who found the missing children # couldn't look after themselves.

Q: No question.

AH: The article revealed the policeman who found the missing children # couldn't look after himself.

Q: No question.

UL: The article revealed the policeman who found that the missing children # couldn't look after themselves.

Q: No question.

UH: The article revealed that the policeman who found the missing children # couldn't look after himself.

Q: No question.

33. AL: The fans recognised the manager who accepted the players # couldn't discipline themselves.

Q: No question.

AH: The fans recognised the manager who accepted the players # couldn't discipline himself.

Q: No question.

UL: The fans recognised the manager who accepted that the players # couldn't discipline themselves.

Q: No question.

UH: The fans recognised that the manager who accepted the players # couldn't discipline himself.

Q: No question.

34. AL: The army accepted the recruits who recognised the enemy commander # couldn't defend himself.

Q: No question.

AH: The army accepted the recruits who recognised the enemy commander # couldn't defend themselves.

Q: No question.

UL: The army accepted the recruits who recognised that the enemy commander # couldn't defend himself.

Q: No question.

UH: The army accepted that the recruits who recognised the enemy commander # couldn't defend themselves.

Q: No question.

35. AL: The workers recognised the company chairman who accepted the payrise # couldn't justify itself.

Q: No question.

AH: The workers recognised the company chairman who accepted the payrise # couldn't justify himself.

Q: No question.

UL: The workers recognised the company chairman who accepted that the payrise # couldn't justify itself.

Q: No question.

UH: The workers recognised that the company chairman who accepted the payrise # couldn't justify himself.

Q: No question.

36. AL: The people accepted the constitution which recognised the dictator # would never establish himself.

Q: No question.

AH: The people accepted the constitution which recognised the dictator # would never establish itself.

Q: No question.

UL: The people accepted the constitution which recognised that the dictator # would never establish himself.

Q: No question.

UH: The people accepted that the constitution which recognised the dictator # would never establish itself.

Q: No question.

A.2 Questionnaire Experiment

The following is the list of materials for the questionnaire study reported as "Experiment 2" in chapter 7. The conditions are to be read as follows: "A" is "Ambiguous", "L" is "Low", and "H" is "High". The two alternative interpretations "a" and "b" are given below each item.

1. A: The investigation revealed the councillors who accepted the spending cuts had not justified themselves.

L: The investigation revealed the councillors who accepted that the spending cuts had not justified themselves.

H: The investigation revealed that the councillors who accepted the spending cuts had not justified themselves.

a. The councillors had not justified themselves.

b. The spending cuts had not justified themselves.

2. A: The children's home accepted the girls who revealed their tormenters had been injecting themselves.

L: The children's home accepted the girls who revealed that their tormenters had been injecting themselves.

H: The children's home accepted that the girls who revealed their tormenters had been injecting themselves.

a. The girls had been injecting themselves.

b. The tormenters had been injecting themselves.

3. A: The editor accepted the stories which revealed the corrupt politicians had contradicted each other.

L: The editor accepted the stories which revealed that the corrupt politicians had contradicted each other.

H: The editor accepted that the stories which revealed the corrupt politicians had contradicted each other.

a. The stories had contradicted each other.

b. The corrupt politicians had contradicted each other.

4. A: The detective forgot the witness who recognised the murderer had disguised himself.

L: The detective forgot the witness who recognised that the murderer had disguised himself.

H: The detective forgot that the witness who recognised the murderer had disguised himself.

a. The witness had disguised himself.

b. The murderer had disguised himself.

5. A: The general recalled the troops who found the enemy spies had shot themselves.

L: The general recalled the troops who found that the enemy spies had shot themselves.

H: The general recalled that the troops who found the enemy spies had shot themselves.

a. The troops had shot themselves.

b. The enemy spies had shot themselves.

6. A: The pensioners recalled the wartime singers who found the German pilots had poisoned themselves.

L: The pensioners recalled the wartime singers who found that the German pilots had poisoned themselves.

H: The pensioners recalled that the wartime singers who found the German pilots had poisoned themselves.

a. The wartime singers had poisoned themselves.
b. The German pilots had poisoned themselves.

7. A: The private detective found the informants who recalled the suspects had disguised themselves.
L: The private detective found the informants who recalled that the suspects had disguised themselves.
H: The private detective found that the informants who recalled the suspects had disguised themselves.

a. The informants had disguised themselves.
b. The suspects had disguised themselves.

8. A: The professor acknowledged the students who understood the new theories had contradicted each other.
L: The professor acknowledged the students who understood that the new theories had contradicted each other.
H: The professor acknowledged that the students who understood the new theories had contradicted each other.

a. The students had contradicted each other.
b. The new theories had contradicted each other.

9. A: The social worker understood the teenager who acknowledged his gang leader couldn't express himself.
L: The social worker understood the teenager who acknowledged that his gang leader couldn't express himself.
H: The social worker understood that the teenager who acknowledged his gang leader couldn't express himself.

a. The teenager couldn't express himself.
b. The gang leader couldn't express himself

10. A: The editor found the photographer who saw the cabinet minister had disgraced himself.
L: The editor found the photographer who saw that the cabinet minister had disgraced himself.
H: The editor found that the photographer who saw the cabinet minister had disgraced himself.

a. The photographer had disgraced himself.
b. The cabinet minister had disgraced himself

11. A: The psychiatrist accepted the male patients who forgot their wives couldn't look after themselves.
L: The psychiatrist accepted the male patients who forgot that their wives couldn't look after themselves.
H: The psychiatrist accepted that the male patients who forgot their wives couldn't look after themselves.

a. The male patients couldn't look after themselves.
b. The wives couldn't look after themselves.

12. A: The inspector recognised the builder who accepted the young apprentice would be a danger to himself.

L: The inspector recognised the builder who accepted that the young apprentice would be a danger to himself.
H: The inspector recognised that the builder who accepted the young apprentice would be a danger to himself.

a. The builder would be a danger to himself.
b. The young apprentice would be a danger to himself.

13. A: The judge accepted the witness who recognised the suspect had been deceiving himself.
L: The judge accepted the witness who recognised that the suspect had been deceiving himself.
H: The judge accepted that the witness who recognised the suspect had been deceiving himself.

a. The witness had been deceiving himself.
b. The suspect had been deceiving himself.

14. A: The tabloids revealed the informer who recalled the old politician had exposed himself.
L: The tabloids revealed the informer who recalled that the old politician had exposed himself.
H: The tabloids revealed that the informer who recalled the old politician had exposed himself.

a. The informer had exposed himself.
b. The old politician had exposed himself.

15. A: The shop steward understood the workers who accepted the new managers could not commit themselves.
L: The shop steward understood the workers who accepted that the new managers could not commit themselves.
H: The shop steward understood that the workers who accepted the new managers could not commit themselves.

a. The workers could not commit themselves.
b. The new managers could not commit themselves,

16. A: The director understood the administrator who accepted the incompetent trainee couldn't explain himself.
L: The director understood the administrator who accepted that the incompetent trainee couldn't explain himself.
H: The director understood that the administrator who accepted the incompetent trainee couldn't explain himself.

a. The administrator couldn't explain himself.
b. The incompetent trainee couldn't explain himself.

17. A: The examiner accepted the students who understood the Latin authors had repeated themselves.
L: The examiner accepted the students who understood that the Latin authors had repeated themselves.
H: The examiner accepted that the students who understood the Latin authors had repeated themselves.

a. The students had repeated themselves.
b. The Latin authors had repeated themselves.

18. A: The personnel officer accepted the applicants who understood the policies had justified themselves.
L: The personnel officer accepted the applicants who understood that the policies had justified themselves.

H: The personnel officer accepted that the applicants who understood the policies had justified themselves.

a. The applicants had justified themselves.
b. The policies had justified themselves.

19. A: The psychoanalyst saw the cult members who recalled the possessed boys couldn't control themselves.

L: The psychoanalyst saw the cult members who recalled that the possessed boys couldn't control themselves.

H: The psychoanalyst saw that the cult members who recalled the possessed boys couldn't control themselves.

a. The possessed boys couldn't control themselves.
b. The cult members couldn't control themselves.

20. A: The camp organiser recalled the volunteers who saw the teenagers had enjoyed themselves.

L: The camp organiser recalled the volunteers who saw that the teenagers had enjoyed themselves.

H: The camp organiser recalled that the volunteers who saw the teenagers had enjoyed themselves.

a. The teenagers had enjoyed themselves.
b. The volunteers had enjoyed themselves.

21. A: The CIA revealed the farmers who saw the aliens had surrendered themselves.

L: The CIA revealed the farmers who saw that the aliens had surrendered themselves.

H: The CIA revealed that the farmers who saw the aliens had surrendered themselves.

a. The aliens had surrendered themselves.
b. The farmers had surrendered themselves.

22. A: The audience saw the band member who acknowledged his promoter had intoxicated himself.

L: The audience saw the band member who acknowledged that his promoter had intoxicated himself.

H: The audience saw that the band member who acknowledged his promoter had intoxicated himself.

a. The promoter had intoxicated himself.
b. The band member had intoxicated himself.

23. A: The publisher recognised the authoress who acknowledged her secretary had disgraced herself.

L: The publisher recognised the authoress who acknowledged that her secretary had disgraced herself.

H: The publisher recognised that the authoress who acknowledged her secretary had disgraced herself.

a. The secretary had disgraced herself.
b. The authoress had disgraced herself.

24. A: The bishop acknowledged the priest who accepted the sinner had reformed himself.

L: The bishop acknowledged the priest who accepted that the sinner had reformed himself.

H: The bishop acknowledged that the priest who accepted the sinner had reformed himself.

a. The sinner had reformed himself.
b. The priest had reformed himself.

25. A: The company boss acknowledged the departments which accepted the young graduates had proved themselves.

L: The company boss acknowledged the departments which accepted that the young graduates had proved themselves.

H: The company boss acknowledged that the departments which accepted the young graduates had proved themselves.

a. The young graduates had proved themselves,
b. The departments had proved themselves.

26. A: The butler saw the prince who revealed his boyfriend had hanged himself.

L: The butler saw the prince who revealed that his boyfriend had hanged himself.

H: The butler saw that the prince who revealed his boyfriend had hanged himself.

a. The boyfriend had hanged himself.
b. The prince had hanged himself.

27. A: The war veterans recalled the officer who revealed the traitor had drowned himself.

L: The war veterans recalled the officer who revealed that the traitor had drowned himself.

H: The war veterans recalled that the officer who revealed the traitor had drowned himself.

a. The traitor had drowned himself.
b. The officer had drowned himself.

28. A: The explorers recalled the tribesman who revealed the witch-doctor had sacrificed himself.

L: The explorers recalled the tribesman who revealed that the witch-doctor had sacrificed himself.

H: The explorers recalled that the tribesman who revealed the witch-doctor had sacrificed himself.

a. The witch-doctor had sacrificed himself.
b. The tribesman had sacrificed himself.

29. A: The scientist found the Eskimo who recalled the famous explorer had no confidence in himself.

L: The scientist found the Eskimo who recalled that the famous explorer had no confidence in himself.

H: The scientist found that the Eskimo who recalled the famous explorer had no confidence in himself.

a. The famous explorer had no confidence in himself.
b. The Eskimo had no confidence in himself.

30. A: The detective found the child who recalled the kidnapper despised himself.

L: The detective found the child who recalled that the kidnapper despised himself.

H: The detective found that the child who recalled the kidnapper despised himself.

a. The kidnapper despised himself.
b. The child despised himself.

31. A: The newspaper report revealed the boys who found the escaped prisoners had injured themselves.

L: The newspaper report revealed the boys who found that the escaped prisoners had injured themselves.

H: The newspaper report revealed that the boys who found the escaped prisoners had injured themselves.

a. The escaped prisoners had injured themselves.
b. The boys had injured themselves.

32. A: The article revealed the policeman who found the missing child couldn't look after himself.

L: The article revealed the policeman who found that the missing child couldn't look after himself.

H: The article revealed that the policeman who found the missing child couldn't look after himself.

a. The missing child couldn't look after himself.
b. The policeman couldn't look after himself.

33. A: The fans recognised the manager who accepted the player couldn't discipline himself.

L: The fans recognised the manager who accepted that the player couldn't discipline himself.

H: The fans recognised that the manager who accepted the player couldn't discipline himself.

a. The player couldn't discipline himself.
b. The manager couldn't discipline himself.

34. A: The army accepted the recruit who recognised the enemy commander couldn't defend himself.

L: The army accepted the recruit who recognised that the enemy commander couldn't defend himself.

H: The army accepted that the recruit who recognised the enemy commander couldn't defend himself.

a. The enemy commander couldn't defend himself.
b. The recruit couldn't defend himself.

35. A: The workers recognised the company directors who accepted the pay increases couldn't justify themselves.

L: The workers recognised the company directors who accepted that the pay increases couldn't justify themselves.

H: The workers recognised that the company directors who accepted the pay increases couldn't justify themselves.

a. The pay increases couldn't justify themselves.
b. The company directors couldn't justify themselves.

36. A: The people accepted the constitution which recognised the dictatorship would never establish itself.

L: The people accepted the constitution which recognised that the dictatorship would never establish itself.

H: The people accepted that the constitution which recognised the dictatorship would never establish itself.

a. The dictatorship would never establish itself.
b. The constitution would never establish itself.

Appendix B

Experimental Materials from chapter 8

B.1 Experiment 1

The following is a list of the materials for the questionnaire study referred to as "Experiment 1" in chapter 8. We show the materials in roman characters, with gloss and English translation, though of course the actual experimental booklets were printed in Japanese script. "TH" indicates the thematic condition (i.e. involving the postposition *kara-no*) and "NT" indicates the non-thematic condition (i.e. involving the postposition *no*). The two possible interpretations appear as "a" and "b" under each item. Items which are shared with Experiment 2 are marked with a star. Items 1-12 are animate-low, and items 13-24 are animate-high.

1*. TH: shokuin ga hometa shichou kara-no kikakusho.
employee NOM praised mayor from plan-document
"the plan document from the mayor that the employees praised"

NT: shokuin ga hometa shichou no kikakusho.
employee NOM praised mayor GEN plan-document
"the plan document of the mayor that the employees praised"

a. shokuin ga shichou wo hometa.
employee NOM mayor ACC praised
"The employees praised the mayor."

b. shokuin ga kikakusho wo hometa.
employee NOM plan-document ACC praised
"The employees praised the plan document."

2. TH: shachou ga hihan-shita maneejaa kara-no messeeji.
managing-director NOM criticised manager from message
"The message from the manager that the managing director criticised."

NT: shachou ga hihan-shita maneejaa no messeeji.
managing-director NOM criticised manager GEN message
"The message of the manager that the managing director criticised."

a. shachou ga maneejaa wo hihan-shita.
managing-director NOM manager ACC criticised.
"The managing director criticised the manager."

b. shachou ga messeeji wo hihan-shita.
managing-director NOM message ACC criticised.
"The managing director criticised the message."

3*. TH: seito ga mushi-sita sensei kara-no chuui.
pupil NOM ignored teacher from warning.
"the warning from the teacher that the pupils ignored"

NT: seito ga mushi-sita sensei no chuui.
pupil NOM ignored teacher GEN warning.
"the warning of the teacher that the pupils ignored"

a. seito ga sensei wo mushi-shita
pupiles NOM teacher ACC ignored
"The pupils ignored the teacher."

b. seito ga chuui wo mushi-shita
pupil NOM warning ACC ignored
"The pupils ignored the warning."

4. TH: chichi ga maishuu matteita sakaya kara-no seikyuusho
father NOM every-week waited wine-merchant from bill
"the bill of the wine merchant that father waited for every week"

NT: chichi ga maishuu matteita sakaya no seikyuusho
father NOM every-week waited wine-merchant GEN bill
"the bill of the wine merchant that father waited for every week"

a. chichi ga maishuu sakaya wo matteita
father NOM every-week wine-merchant ACC waited
"Father waited for the wine merchant every week."

b. chichi ga maishuu seikyuusho wo matteita
father NOM every-week bill ACC waited
"Father waited for the bill every week."

5. TH: kyoushi ga kowagaru kouchousensei kara-no iitsuke
teachers NOM fear head-teacher from instructions
"the instructions from the head teacher that the teachers feared.

NT: kyoushi ga kowagaru kouchousensei no iitsuke
teachers NOM fear head-teacher GEN instructions
"the instructions of the head teacher that the teachers feared.

a. kyoushi ga kouchousensei wo kowagaru.
teachers NOM head-teacher ACC fear.
"The teachers fear the head teacher."

b. kyoushi ga iitsuke wo kowagaru.
teachers NOM instruction ACC fear.
"The teachers fear the instruction."

6*. TH: obaasan ga natsukashiku omoidasu doukyuusei kara-no yosegaki
granny NOM nostalgically remember class-mates from leaving-card
"the leaving card from the classmates that Granny nostalgically remembers"

NT: obaasan ga natsukashiku omoidasu doukyuusei no yosegaki
granny NOM nostalgically remember class-mates GEN leaving-card
"the leaving card of the classmates that Granny nostalgically remembers"

a. obaasan ga dioukyuusei wo natsukasiku omoidasu.
granny NOM classmates ACC nostalgically remembers
"Granny nostalgically remembers the classmates."

b. obaasan ga yosegaki wo natsukasiku omoidasu.
granny NOM leaving-card ACC nostalgically remembers
"Granny nostalgically remembers the leaving card."

7*. TH: yukiko ga gakkou de mitsuketa koibito kara-no tegami
Yukiko NOM school LOC found boyfriend from letter
"The letter from the boyfriend that Yukiko found at school"

NT: yukiko ga gakkou de mitsuketa koibito no tegami
Yukiko NOM school LOC found boyfriend from letter
"The letter of the boyfriend that Yukiko found at school"

a. Yukiko ga gakkou de tegami wo mitsuketa.
Yukiko NOM school LOC letter ACC found
"Yukiko found the letter at school."

b. Yukiko ga gakkou de koibito wo mitsuketa.
Yukiko NOM school LOC boyfriend ACC found
"Yukiko found (met) the boyfriend at school."

8*. TH: risaachaa ga hinan shita kagakusha kara-no houkokusho
researcher NOM criticised scientist from report
"the report from the scientist that the researcher criticised"

NT: risaachaa ga hinan shita kagakusha no houkokusho
researcher NOM criticised scientist GEN report
"the report of the scientist that the researcher criticised"

a. risaachaa ga houkokusho wo hinan shita
researcher NOM report ACC criticised
"The researcher criticised the report."

b. risaachaa ga kagakusha wo hinan shita
researcher NOM scientist ACC criticised
"The researcher criticised the scientist."

9*. TH: kodomotachi ga wasureteita ojisan kara-no otoshidama
children NOM had forgotten uncle from new year's present
"the new year's present from the uncle that the children had forgotten"

NT: kodomotachi ga wasureteita ojisan no otoshidama
children NOM had forgotten uncle GEN new year's present
"the new year's present of the uncle that the children had forgotten"

a. kodomotachi ga otoshidama no koto wo wasureteita
children NOM new year's present GEN fact ACC had forgotten
"The children had forgotten (about) the new year's present."

b. kodomotachi ga otoshidama no koto wo wasureteita
children NOM new year's present GEN fact ACC had forgotten
"The children had forgotten (about) the uncle."

10. TH: hanji ga azawaratta bengoshi kara-no shorui
judge NOM laughed sardonically lawyer from document
"the documents from the lawyer that the judge laughed at sardonically"

NT: hanji ga azawaratta bengoshi no shorui
judge NOM laughed sardonically lawyer GEN document
"the documents of the lawyer that the judge laughed at sardonically"

a. hanji ga shorui wo azawaratta.
judge NOM document ACC laughed sardonically
"The judge laughed sardonically at/about the documents."

b. hanji ga shorui wo azawaratta.
judge NOM document ACC laughed sardonically
"The judge laughed sardonically at the lawyer."

11*. TH: hanji ga shinjita mokugekisha kara-no repooto
judge NOM believed witness from report.
"the report from the witness that the judge believed"

NT: hanji ga shinjita mokugekisha no repooto
judge NOM believed witness GEN report
"the report of the witness that the judge believed"

a. hanji ga repooto wo shinjita
judge NOM report ACC believed
"The judge believed the report."

b. hanji ga mokugekisha wo shinjita
judge NOM witness ACC believed
"The judge believed the witness."

12*. TH: kachou ga oboeteita buka kara-no memo
section-chief NOM remembered subordinate from memo
"the memo from the subordinate that the section chief remembered"

NT: kachou ga oboeteita buka no memo
section-chief NOM remembered subordinate GEN memo
"the memo of the subordinate that the section chief remembered"

a. kachou ga memo wo oboeteita.
section-chief NOM memo ACC remembered
"The section chief remembered the memo."

b. kachou ga buka wo oboeteita.
section-chief NOM subordinate ACC remembered
"The section chief remembered the subordinate."

13. TH: chichi ga shitteiru gyoson kara-no gakusei
father NOM knows fishing-village from student
"the student from the fishing village that father knows"

NT: chichi ga shitteiru gyoson no gakusei
father NOM knows fishing-village of student
"the student of the fishing village that father knows"

a. chichi ga gyoson wo sitteiru.
father NOM fishing-village ACC knows
"Father knows the fishing village."

b. chichi ga gakesei wo sitteiru.
father NOM student ACC knows
"Father knows the student."

14. TH: gekai ga shujutsu-shita byouin kara-no kanja
surgeon NOM operated hospital from patient
"the patient from the hospital that/where the surgeon operated (on)"

NT: gekai ga shujutsu-shita byouin no kanja
surgeon NOM operated hospital of patient
"the patient of the hospital that/where the surgeon operated (on)"

a. gekai ga byouin de shujutsu-shita.
suegeon NOM hospital LOC operated
"The surgeon operated at the hospital."

b. gekai ga kanja wo shujutsu-shita.
suegeon NOM patient ACC operated
"The surgeon operated at the hospital."

15*. TH: kankoukyaku ga shashin wo totta shima kara-no ryoushi
tourists NOM photo ACC took island from fisherman
"the fisherman from the island that the tourists took pictures of"

NT: kankoukyaku ga shashin wo totta shima no ryoushi
tourists NOM photo ACC took island GEN fisherman
"the fisherman of the island that the tourists took pictures of"

a. kankoukyaku ga shima no shashin wo totta.
tourists NOM island GEN photo ACC took
"The tourists took photos of the island."

b. kankoukyaku ga ryoushi no shashin wo totta.
tourists NOM fisherman GEN photo ACC took
"The tourists took photos of the fisherman."

16*. TH: zasshi ga shoukai-shita tarentoshoo kara-no geinoujin
magazine NOM introduced talent-show from show-biz-personality
"the show business personality from the talent show that the magazine introduced"

NT: zasshi ga shoukai-shita tarentoshoo no geinoujin
magazine NOM introduced talent-show GEN show-biz-personality
"the show business personality of the talent show that the magazine introduced"

a. zasshi ga tarentoshoo wo shoukai-shita
magazine NOM talent-show ACC introduced
"The magazine introduced the talent show."

b. zasshi ga geinoujin wo shoukai-shita
magazine NOM show-biz-personality ACC introduced
"The magazine introduced the show business personality."

17*. TH: terebikyoku ga satsuei-shita sumoubeya kara-no rikishi
TV-company NOM filmed Sumo-stable from Sumo-wrestler
"the Sumo wrestler from the stable that the TV company filmed"

NT: terebikyoku ga satsuei-shita sumoubeya no rikishi
TV-company NOM filmed Sumo-stable GEN Sumo-wrestler
"the Sumo wrestler of the stable that the TV company filmed"

a. terebikyoku ga sumoubeya wo satsuei-shita.
TV-company NOM Sumo-stable ACC filmed
"The TV company filmed the Sumo-stable."

b. terebikyoku ga rikishi wo satsuei-shita.
TV-company NOM Sumo-wrestler ACC filmed
"The TV company filmed the Sumo wrestler."

18. TH: shufu ga shin'you-shiteita takushiigaisha kara-no untenshu
housewives NOM trusted taxi-company from driver
"the driver from the taxi company that the housewives trusted"

NT: shufu ga shin'you-shiteita takushiigaisha no untenshu
housewives NOM trusted taxi-company GEN driver
"the driver of the taxi company that the housewives trusted"

a. shufu ga takushiigaisha wo shin'you-shiteita.
housewives NOM taxi-company ACC trusted
"The housewives trusted the taxi company."

b. shufu ga unteshu wo shin'you-shiteita.
housewives NOM driver ACC trusted
"The housewives trusted the driver."

19. TH: daimyou ga sagashiteita oshiro kara-no ninja
feudal-lord NOM was-looking-for castle from ninja
"the ninja from the castle that the feudal lord was looking for"

NT: daimyou ga sagashiteita oshiro no ninja
feudal-lord NOM was-looking-for castle GEN ninja
"the ninja of the castle that the feudal lord was looking for"

a. daimyou ga ninja wo sagashiteita.
feudal-lord NOM ninja ACC was-looking-for
"The feudal lord was looking for the ninja."

b. daimyou ga oshiro wo sagashiteita.
feudal-lord NOM castle ACC was-looking-for
"The feudal lord was looking for the castle."

20. TH: kodomotachi ga egaita yama kara-no saru
children NOM painted mountain from monkey
"the monkey from the mountain that the children painted"

NT: kodomotachi ga egaita yama no saru
children NOM painted mountain GEN monkey
"the monkey of the mountain that the children painted"

a. kodomotachi ga saru wo egita.
children NOM monkey ACC painted
"The children painted a picture of the monkey."

b. kodomotachi ga yama wo egita.
children NOM mountain ACC painted
"The children painted a picture of the mountain."

21. TH: minkanjin ga osoreta guntai kara-no shireikan
civilians NOM feared army from commander
"the commander from the army that the civilians feared"

NT: minkanjin ga osoreta guntai no shireikan
civilians NOM feared army GEN commander
"the commander of the army that the civilians feared"

a. minkanjin ga shireikan wo osoreta.
civilians NOM commander ACC feared
"The civilians feared the commander."

b. minkanjin ga guntai wo osoreta.
civilians NOM army ACC feared
"The civilians feared the army."

22*. TH: genjuumin ga odoshita tankentai kara-no taichou
natives NOM threatened expedition-force from commander
"the commander from the expedition force that the natives threatened"

NT: genjuumin ga odoshita tankentai no taichou
natives NOM threatened expedition-force GEN commander
"the commander of the expedition force that the natives threatened"

a. genjuumin ga taichou wo odoshita.
natives NOM commander ACC threatened
"The natives threatened the commander."

b. genjuumin ga tankentai wo odoshita.
natives NOM expedition-force ACC threatened
"The natives threatened the expedition force."

23. TH: heishi ga mamotta nanminkyampu kara-no kyoushi
soldiers NOM defended refugee-camp from teacher
"the teacher from the refugee camp that the soldiers defended"

NT: heishi ga mamotta nanminkyampu no kyoushi
soldiers NOM defended refugee-camp GEN teacher
"the teacher of the refugee camp that the soldiers defended"

a. heishi ga kyoushi wo mamotta.
soldiers NOM teacher ACC defended
"The soldiers defended the teacher."

b. heishi ga nanminkyampu wo mamotta.
soldiers NOM refugee camp ACC defended
"The soldiers defended the refugee camp."

24*. TH: keisatsu ga utagatteita kyoudan kara-no dendoushi
police NOM suspected cult from missionary
"the missionary from the cult that the police suspected"

NT: keisatsu ga utagatteita kyoudan no dendoushi
police NOM suspected cult GEN missionary
"the missionary of the cult that the police suspected"

a. keisatsu ga dendoushi wo utagatteita.
police NOM missionary ACC suspected
"The police suspected the missionary."

b. keisatsu ga kyoudan wo utagatteita.
police NOM cult ACC suspected
"The police suspected the cult."

B.2 Experiment 2

The following is a list of the materials for the questionnaire study referred to as "Experiment 2" in chapter 8. We show the materials in roman characters, with gloss and English translation, though of course the actual experimental booklets were printed in Japanese script. "CL" indicates the clausal condition (i.e. the condition which includes a clause boundary between the two possible attachment sites). "TH" indicates the thematic condition (i.e. involving the postposition *kara-no*) and "NT" indicates the non-thematic condition (i.e. involving the postposition *no*). The two possible interpretations appear as "a" and "b" under each item. We do not repeat text which already appears in appendix B.1 in the materials for Experiment 1. Items which are shared with Experiment 1 are marked with a star. Items 1-12 are animate-low, and items 13-24 are animate-high.

1*. CL: yukiko ga gakkou de mitsuketa koibito ga kaita tegami
Yukiko NOM school LOC found boyfriend NOM wrote letter
"the letter that the boyfriend wrote that Yukiko found/met at school"
(for TH, NT and interpretations, see Experiment 1, item 7)

2*. CL: obaasan ga omoidashita doukyuusei ga kaita yosegaki
Granny NOM remembered classmates NOM wrote leaving-card
"the leaving card that the classmates wrote that Granny remembered"

TH: obaasan ga omoidashita doukyuusei kara-no yosegaki
Granny NOM remembered classmates from leaving-card
"the leaving card from the classmates that Granny remembered"

NT: obaasan ga omoidashita doukyuusei no yosegaki
Granny NOM remembered classmates GEN leaving-card
"the leaving card of the classmates that Granny remembered"
(for interpretations, see Experiment 1, item 6 [1])

3.* CL: shokuin ga hometa shichou ga kaita kikakusho.
employee NOM praised mayor NOM wrote plan-document
"the plan document that the mayor wrote that the employee praised"
(for TH, NT and interpretations, see Experiment 1, item 1)

4.* CL: seito ga mushi-shita sensei ga shita chuui
students NOM ignored teacher NOM did warning
"the warning that the teacher made that the students ignored"
(for TH, NT and interpretations, see Experiment 1, item 3)

5.* CL: hanji ga shinjita mokugekisha ga sakusei-shita repooto
judge NOM believed witness NOM drew-up report
"the report that the witness drew up that the judge believed"
(for TH, NT and interpretations, see Experiment 1, item 11)

6.* CL: risaachaa ga hihan-shita kagakusha ga dashita houkokusho
researcher NOM criticized scientist NOM sent report
"the report that the scientist sent that the researcher criticised"
(for TH, NT and interpretations, see Experiment 1, item 8)

7.* CL: kodomotachi ga wasureteita ojisan ga kureta otoshidama
children NOM forgot uncle NOM gave new-year-present
"the new year present that the uncle gave that the children had forgotten (about)"
(for TH, NT and interpretations, see Experiment 1, item 9)

8. CL: heishi ga mushi-shita chuusa ga sakenda shirei
soldiers NOM ignored lieutentant NOM shouted command
"the command that the lieutentant sakenda that the soldiers ignored"

TH: heishi ga mushi-shita chuusa kara-no shirei
soldiers NOM ignored lieutentant from command
"the command from the lietenant that the soldiers ignored"

NT: heishi ga mushi-shita chuusa no shirei
soldiers NOM ignored lieutentant GEN command

a. heishi ga shirei wo mushi-shita.
soldiers NOM command ACC ignored
"The soldiers ignored the command."

b. heishi ga chuusa wo mushi-shita.
soldiers NOM lieutenant ACC ignored
"The soldiers ignored the lieutenant."

9. CL: yakunin ga okotta seerusuman ga watashita wairo
official NOM got-angry salesman NOM passed bribe
"the bribe that the salesman passed that the official got angry about"

[1] Although this material is shared with Experiment 1, the wording of the relative clause is slightly different, so we show the wording of all conditions.

TH: yakunin ga okotta seerusuman kara-no wairo
official NOM got-angry salesman from bribe
"the bribe from the salesman that the official got angry about"

NT: yakunin ga okotta seerusuman no wairo
official NOM got-angry salesman GEN bribe

a. yakunin ga wairo ni okotta.
official NOM bribe DAT got-angry
"The official got angry about the bribe."

b. yakunin ga seerusuman ni okotta.
official NOM salesman DAT got-angry
"The official got angry with the salesman."

10. CL: hisho ga mitsukerarenakatta buchou ga yonda fakkusu
secretary NOM could-not-find department-chief NOM read fax
"the fax that the department chief read that the secretary couldn't find"

TH: hisho ga mitsukerarenakatta buchou kara-no fakkusu
secretary NOM could-not-find department-chief from fax
"the fax from the department chief that the secretary couldn't find"

NT: hisho ga mitsukerarenakatta buchou no fakkusu
secretary NOM could-not-find department-chief GEN fax
"the fax of the department chief that the secretary couldn't find"

a. hisho ga buchou wo mitsukerarenakatta.
secretary NOM department-chief ACC could-not-find
"The secretary couldn't find the department chief."

b. hisho ga fakkusu wo mitsukerarenakatta.
secretary NOM fax ACC could-not-find
"The secretary couldn't find the fax."

11.* CL: kachou ga oboeteita buka ga hatta memo
section-chief NOM remembered subordinate NOM hatta memo
"the memo that the subordinate stuck (e.g. on the notice board) that the section chief remembered"
(for TH, NT and interpretations, see experiment 1, item 12)

12. CL: roudousha ga goui-shita chouteisha ga kangaeta teian
worker NOM agreed arbitrator NOM thought proposal
"the proposal which the arbitrator thought of that the workers agreed with"

TH: roudousha ga goui-shita chouteisha kara-no teian
worker NOM agreed arbitrator from proposal
"the proposal from the arbitrator that the workers agreed with"

NT: roudousha ga goui-shita chouteisha no teian
worker NOM agreed arbitrator gen proposal
"the proposal of the arbitrator that the workers agreed with

a. roudousha ga chouteisha ni goui-shita
workers NOM arbitrator DAT agreed
"The workers agreed with the arbitrator."

b. roudousha ga teian ni goui-shita
workers NOM proposal DAT agreed
"The workers agreed with the proposal."

13. CL: heishi ga osotta kyuuden kara kita shisha
soldiers NOM attacked palace from came messanger
"the messanger that came from the palace that the soldiers attacked"

TH: heishi ga osotta kyuuden kara-no shisha
soldiers NOM attacked palace from messanger
"the messanger from the palace that the soldiers attacked"

NT: heishi ga osotta kyuuden no shisha
soldiers NOM attacked palace GEN messanger
"the messanger of the palace that the soldiers attacked"

a. heishi ga kyuuden wo osotta.
soldiers NOM palace ACC attacked
"The soldiers attacked the palace."

b. heishi ga shisha wo osotta.
soldiers NOM messanger ACC attacked
"The soldiers attacked the messanger."

14. CL: junreisha ga houmon-shita seichi kara kita seishokusha
pilgrims NOM visited holy-place from came priest
"the priest that came from the holy place that the pilgrims visited"

TH: junreisha ga houmon-shita seichi kara-no seishokusha
pilgrims NOM visited holy-place from priest
"the priest from the holy place that the pilgrims visited"

NT: junreisha ga houmon-shita seichi no seishokusha
pilgrims NOM visited holy-place of priest
"the priest of the holy place that the pilgrims visited"

a. junreisha ga seishokusha wo otozureta.
pilgrims NOM priest ACC visited
"The pilgrims visited the priest."

b. junreisha ga seichi wo otozureta.
pilgrims NOM holy-place ACC visited
"The pilgrims visited the holy-place."

15.* CL: kankoukyaku ga shashin wo totta shima kara kita ryoushi
tourists NOM photo ACC took island from came fisherman
"the fisherman that came from the island that the tourists took pictures of"

(for TH, NT and interpretations, see Experiment 1, item 15)

16. CL: kokuminn ga eranda seitou wo daihyou-suru kouhosha
citizens NOM chose party ACC represent candidate
"the candidate who represented the party that the citizens chose"

TH: kokuminn ga eranda seitou kara-no kouhosha
citizens NOM chose party from candidate
"the candidate from the party that the citizens chose"

NT: kokuminn ga eranda seitou no kouhosha
citizens NOM chose party GEN candidate
"the candidate of the party that the citizens chose"

a. kokumin ga senkyo de sono seitou wo eranda
citizens NOM election LOC that party ACC chose
"The citizens chose the party at the election."

b. kokumin ga senkyo de sono kouhosha wo eranda
citizens NOM election LOC that candidate ACC chose
"The citizens chose the candidate at the election."

17.* CL: zasshi ga shoukai-shita tarentoshoo ni deta geinoujin
magazine NOM introduced talent-show DAT appeared show-biz-personality
"the show business personality who appeared in the talent show that the magazine introduced"

(for TH, NT and interpretations, see Experiment 1, item 16)

18.* CL: terebikyoku ga satsuei-shita sumoubeya ni haitta rikishi
TV-company NOM filmed sumo-stable DAT entered sumo-wrestler
"the Sumo wrestler who entered the stable that the TV company filmed"

(for TH, NT and interpretations, see Experiment 1, item 17)

19. CL: shiyakusho ga keiyaku-shita kensetsugaisha ni nyuusha-shita sekkeisha
city-hall NOM contracted construction-company DAT joined-company designer
"the designer who joined the construction company that the city hall contracted"

TH: shiyakusho ga keiyaku-shita kensetsugaisha kara-no sekkeisha
city-hall NOM contracted construction-company from designer
"the designer from the construction company that the city hall contracted"

NT: shiyakusho ga keiyaku-shita kensetsugaisha no sekkeisha
city-hall NOM contracted construction-company GEN designer
"the designer of the construction company that the city hall contracted"

a. shiyakusho ga sekkeisha to keiyaku-shita
city-hall NOM designer with contracted
"The city hall contracted the designer."

b. shiyakusho ga kensetsugaisha to keiyaku-shita
city-hall NOM construction-company with contracted
"The city hall contracted the construction company."

20.* CL: keisatsu ga utagatteita kyoudan wo tsukutta dendoushi
police NOM suspected cult ACC founded missionary
"the missionary who founded the cult that the police suspeced"

(for THG, NT and interpretations, please see Experiment 1, item 24)

21 CL: boudousha ga osotta keisatsusho wo mamotta keikan
rioter NOM attacked police-station ACC defended policeman
"the policeman that defended the police station that the rioter attacked"

TH: boudousha ga osotta keisatsusho kara-no keikan
rioter NOM attacked police-station from policeman
"the policeman from the police station that the rioters attacked"

NT: boudousha ga osotta keisatsusho no keikan
rioter NOM attacked police-station GEN policeman
"the policeman of the police station that the rioters attacked"

a. bouousha ga keisatsusho wo osotta.
rioter NOM police-station ACC attacked
"The rioters attacked the police station."

b. bouousha ga keikan wo osotta.
rioter NOM policeman ACC attacked
"The rioters attacked the policeman."

22. CL: kangofu ga denwa-shita byouin ni ita isha
nurse NOM phoned hospital DAT was doctor
"the doctor who was in the hospital that the nurse phoned"

TH: kangofu ga denwa-shita byouin kara-no isha
nurse NOM phoned hospital from doctor
"the doctor from the hospital that the nurse phoned"

NT: kangofu ga denwa-shita byouin no isha
nurse NOM phoned hospital GEN doctor
"the doctor of the hospital that the nurse phoned"

a. kangofu ga byouin ni denwa-shita.
nurse NOM hospital DAT phoned
"The nurse phoned the hospital."

b. kangofu ga isha ni denwa-shita.
nurse NOM doctor DAT phoned
"The nurse phoned the doctor."

23.* CL: genjuumin ga odoshita tankentai wo hikiita taichou
natives NOM threatened expedition-force ACC led commander
"the commander who led the expedition force that the natives threatened"

(for TH, NT and interpretation, see Experiment 1, item 22)

24. CL: geijutsuka ga egaita tankou wo nigeta koufu
artist NOM painted coal-mine ACC escaped miner
"the miner who escaped from the coal mine that the artist painted a picture of"

TH: geijutsuka ga egaita tankou kara-no koufu
artist NOM painted coal-mine from miner
"the miner from the coal mine that the artist painted

NT: geijutsuka ga egaita tankou no koufu
artist NOM painted coal-mine GEN miner
"the miner of the coal mine that the artist painted a picture of"

a. geijutsuka ga tankou wo egaita.
artist NOM coal-mine ACC pained.
"The artist painted a picture of the coal mine."

b. geijutsuka ga koufu wo egaita.
artist NOM miner ACC pained.
"The artist painted a picture of the miner."

References

Abney, S. P. 1987. Licensing and parsing. In *Proceedings of NELS*, pages 1–15, University of Massachusetts, Amherst.

Abney, S. P. 1989. A computational model of human parsing. *Journal of Psycholinguistic Research*, 18(1):129–144.

Abney, S. P. and M. Johnson. 1991. Memory requirements and local ambiguities of parsing strategies. *Journal of Psycholinguistic Research*, 20(3):233–250.

Adams, B. C. 1995. A model for strategic reanalysis in sentence processing. Manuscript, University of Virginia.

Altmann, G. and M. Steedman. 1988. Interaction with context during human sentence processing. *Cognition*, 30.

Baccino, T., M. De Vincenzi, and R. Job. 1996. Cross-linguistic studies of the late closure strategy: French and italian. Paper presented at the *Architectures and Mechanisms for Language Processing* conference, Torino, Italy, September.

Bader, M. 1996. On reanalysis: Evidence from German. Manuscript, Friedrich-Schiller-Universität Jena.

Bader, M. and I. Lasser. 1994. German verb-final clauses and sentence processing. In C. Clifton, L. Frazier, and K. Rayner, editors, *Perspectives on Sentence Processing*. Lawrence Erlbaum Associates, New Jersey.

Barker, C. and G. K. Pullum. 1990. A theory of command relations. *Linguistics and Philosophy*, 13(1):1–34.

Barton, E. and R. Berwick. 1985. Parsing with assertion sets and informational monotonicity. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, pages 769–771, University of California, Los Angeles, August.

Berwick, R. and A. Weinberg. 1984. *The Grammatical Basis of Linguistic Performance: Language Use and Acquisition*. MIT Press, Cambridge: MA.

Branigan, H., P. Sturt, and Y. Matsumoto-Sturt. 1996. Left branching attachment and thematic domains. Poster presented at the *Architectures and Mechanisms for Language Processing* conference, Torino, Italy, September.

Brysbaert, M. and D.C. Mitchell. 1996. Modifier attachment in dutch: Deciding between garden-path, construal and statistical tuning accounts of parsing. Paper presented at the workshop on Computational Models of Human Syntactic Processing held at NIAS, Wassenaar, June 1996.

Cho, Young-mee Yu and P. Sells. 1995. A lexical account of inflectional suffixes in Korean. *Journal of East Asian Linguistics*, 4:119–174.

Chomsky, N. 1981. *Lectures on Government and Binding*. Foris.

Chomsky, N. 1986a. *Barriers*. MIT Press, Cambridge, MA.

Chomsky, N. 1986b. *Knowledge of language: its nature, origin and use*. Praeger, New York.

Chomsky, N. 1992. A minimalist program for linguistic theory. Occasional Papers in Linguistics, 1; MIT, Cambridge, MA.

Clifton, C. 1995. Evaluating models of human sentence processing. Paper presented at *Architectures and Mechanisms for Language Processing* Conference, Edinburgh, 1995, December. (to appear in the conference proceedings).

Clifton, C., S. Speer, and S.P. Abney. 1991. Parsing arguments: Phrase structure and argument structure as determinants of initial parsing decisions. *Journal of Memory and Language*, 30:251–271.

Cornell, T.L. 1994. On determining the consistency of partial descriptions of trees. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, La Cruces, New Mexico, June.

Crain, S. and M. Steedman. 1985. On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Karttunen, and A. Zwicky, editors, *Natural Language Processing: Psychological, Computational and Theoretical Perspectives*. Cambridge University Press, Cambridge, chapter 10, pages 320–358.

Crocker, M.W. 1994. On the nature of the principle-based sentence processor. In C. Clifton, L. Frazier, and K. Rayner, editors, *Perspectives on Sentence Processing*. Lawrence Erlbaum Associates, New Jersey, chapter 11, pages 245–266.

Crocker, M.W. 1996. *Computational Psycholinguistics: An Interdisciplinary Approach to the Study of Language*. Studies in Theoretical Psycholinguistics 20. Kluwer Academic Publishers, The Netherlands.

Cuetos, F. and D.C. Mitchell. 1988. Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition*, 30:72–105.

De Vincenzi, M. and R. Job. 1993. Some observations on the universality of the late-closure strategy. *Journal of Psycholinguistic Research*, 22(2):189–206.

De Vincenzi, M. and R. Job. 1995. An investigation of late closure: the role of syntax, thematic structure and pragmatics in initial and final interpretation. *Journal of Experimental Psychology. Learning, Memory and Cognition*, 21(5):1303–1321.

Dowty, D. 1988. Type raising, functional composition, and non-constituent conjunction. In *Categorial Grammars and Natural Language Structures*, Studies in Linguistics and Philosophy, Volume 32. D. Reidel, Dordrecht, pages 153–197.

Ferreira, F. and J.M. Henderson. 1990. Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 16(4):725–745.

Ferreira, F. and J.M. Henderson. 1991. Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, 30:725–745.

Fodor, J.D. and A. Inoue. 1994. The diagnosis and cure of garden paths. *Journal of Psycholinguistic Research*, 23(4):405–432.

Fodor, J.D. and A. Inoue. to appear. Attach Anyway. to appear in J.D. Fodor and F. Ferreira, editors, *Reanalysis in Sentence Processing*, Kluwer Academic Publishers.

Frank, R. 1990. Computation and linguistic theory: A government and binding parser using tree adjoining grammar. Technical Report MS-CIS-90-29, Department of Computer and Information Science, University of Pennsylvania, PA.

Frank, R. 1992. Syntactic locality and tree adjoining grammar: grammatical, acquisition and processing perspectives. IRCS Report 92-47, Department of Computer and Information Science, University of Pennsylvania, PA". PhD Thesis.

Frazier, L. 1978. *On comprehending sentences: Syntactic parsing strategies*. Ph.D. thesis, University of Connecticut.

Frazier, L. 1987. Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory*, 5(4):519–559.

Frazier, L. 1990a. Identifying structure under X^0. *Yearbook of morphology*, 3:87–105.

Frazier, L. 1990b. Parsing modifiers: Special purpose routines in the HPSM? In D.A. Balota, G.B. Flores d'Arcais, and K. Rayner, editors, *Comprehension processes in reading*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, pages 303–331.

Frazier, L. 1994. Sentence (re)-analysis. Paper presented at the *7th Annual CUNY Conference on Human Sentence Processing*, New York, March.

Frazier, L. and C. Clifton. 1996. *Construal*. MIT Press, Cambridge MA.

Frazier, L. and C. Clifton. to appear. Sentence reanalysis, and visibility. to appear in J.D. Fodor and F. Ferreira, editors, *Reanalysis in Sentence Processing*, Kluwer Academic Publishers.

Frazier, L. and J.D. Fodor. 1978. The sausage machine: A new two-stage parsing model. *Cognition*, 6:291–325.

Frazier, L. and K. Rayner. 1982. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14:178–210.

Frazier, L. and K. Rayner. 1987. Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences. *Journal of Memory and Language*, 26:505–526.

Frazier, L. and K. Rayner. 1988. Parameterizing the language processing system: Left-vs. right-branching within and across languages. In J.A. Hawkins, editor, *Explaining Language Universals*. Basil Blackwell, pages 247–279.

Fukui, N. 1986. *A Theory of Category Projections and its Applications*. Ph.D. thesis, MIT.

Gazdar, G. and C. Mellish. 1989. *Natural Language Processing in* PROLOG*: An Introduction to Computational Linguistics*. Addison-Wesley.

Gibson, E. 1991. *A Computational Theory of Human Linguistic Processing: Memory Limitations and Processing breakdown*. Ph.D. thesis, Carnegie Mellon University.

Gibson, E., N. Pearlmutter, E. Canesco-Gonzalez, and G. Hickok. 1996. Recency preference in the human sentence processing mechanism. *Cognition*, 59(1):23–59.

Gilboy, E., J.M. Sopena, C. Clifton, Jr, and L. Frazier. 1995. Argument structure and association preferences in Spanish and English compound NPs. *Cognition*, 54:131–167.

Gorrell, P. 1995a. Japanese trees and the garden path. In R. Mazuka and N. Nagai, editors, *Japanese Sentence Processing*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, chapter 14, pages 331–350.

Gorrell, P. 1995b. *Syntax and Parsing*. Cambridge University Press, Cambridge.

Hemforth, B., L. Konieczny, and C. Scheepers. 1994a. On reanalysis: Head position and syntactic complexity. In B. Hemforth, L. Konieczny, C. Scheepers, and G. Strube, editors, *IIG-Berichte 8/94*. Institut für Informatik und Gesellschaft, IIG, University of Freiburg, pages 23–50.

Hemforth, B., L. Konieczny, and C. Scheepers. 1994b. Principle-based or probabilistic approaches to human parsing. In B. Hemforth, L. Konieczny, C. Scheepers, and G. Strube, editors, *IIG-Berichte 8/94*. Institut für Informatik und Gesellschaft, IIG, University of Freiburg, pages 77–88.

Hemforth, B., L. Konieczny, and C. Scheepers. 1995. Syntactic attachment and anaphor resolution: The two sides of relative clause attachment. Paper presented at the *Architectures and Mechanisms for Language Processing* conference, Edinburgh, December.

Hirose, Y. 1997. Head noun ambiguity and relative clause reanalysis in japanese. Paper presented at the *10th Annual CUNY Conference on Human Sentence Processing*, University of Southern California, March.

Holmes, V.M., L. Stowe, and L. Cupples. 1989. Lexical expectations in parsing complement verb sentences. *Journal of Memory and Language*, 28:668–689.

Hudson, R. 1990. *English Word Grammar*. Blackwell.

Hudson, R. 1994. *Word Grammar*. Blackwell.

Inoue, A. 1991. *A comparative study of parsing in English and Japanese*. Ph.D. thesis, University of Conneticut, Storrs CT.

Inoue, A. and J. Fodor. 1995. Information-paced parsing of Japanese. In R. Mazuka and N. Nagai, editors, *Japanese Sentence Processing*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, chapter 2, pages 9–63.

Joshi, A.K., L.S. Levy, and M. Takahashi. 1975. Tree adjunct grammars. *Journal of Computer and System Sciences*, 10:136–163.

Joshi, A.K., K. Vijay-Shanker, and D. Weir. 1991. The convergence of mildly context-sensitive grammar formalisms. In P. Sells, S.M. Shieber, and T. Wasow, editors, *Foundational Issues in Natural Language Processing*. MIT Press, Cambridge, MA, chapter 2, pages 31–81.

Just, M., P. Carpenter, and J. Wooley. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111:228–238.

Kamide, Y. and D.C. Mitchell. 1996. Relative clause attachment: Evidence from Japanese. Poster presented at the 9th Annual Conference on Human Sentence Processing, New York.

Kimball, J. 1973. Seven principles of surface structure parsing in natural language. *Cognition*, 2(1):15–47.

Konieczny, L. 1996. *Human Sentence Processing: A Semantics-Oriented Approach*. Ph.D. thesis, University of Freiburg.

Konieczny, L., C. Scheepers, and B. Hemforth. 1994. Reanalyses vs. internal repairs: non-monotonic processes in sentence preception. In B. Hemforth, L. Konieczny, C. Scheepers, and G. Strube, editors, *IIG-Berichte 8/94*. Institut für Informatik und Gesellschaft, IIG, University of Freiburg, pages 2–22.

Kuroda, S-Y. 1988. Whether we agree or not: a comparative syntax of English and Japanese. In *Papers from the Second Workshop on Japanese Syntax*. CSLI Publications, Stanford, CA, pages 103–143.

Lambek, J. 1958. The mathematics of sentence structure. *American Mathematical Monthly*, 65:154–170.

Latecki, L. 1991. An indexing technique for implementing command relations. In *Proceedings of the 5th Conference of the European Chapter of the Association for Computational Linguistics*, pages 51–56.

Lewis, R. 1993. *An Architecturally-based Theory of Human Sentence Comprehension*. Ph.D. thesis, Computer Science Department, Carnegie Mellon University.

Lombardo, V. to appear. A computational model of recovery. to appear in J.D. Fodor and F. Ferreira, editors, *Reanalysis in Sentence Processing*, Kluwer Academic Publishers.

Lombardo, V. and P. Sturt. 1997. Incremental processing and infinite local ambiguity. In *Proceedings of the 19th Annual Conference of the Cognitive Science Society. Stanford*.

MacDonald, M.C., N.J. Pearlmutter, and M.S. Seidenberg. 1994. Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4):676–703.

Marantz, A. 1984. *On the Nature of Grammatical Relations*. MIT Press, Cambridge, MA.

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

Marcus, M., D. Hindle, and M. Fleck. 1983. D-theory: Talking about talking about trees. In *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, pages 129–136, Cambridge, MA, june.

Marr, D. 1982. *Vision: A computational investigation into the human representation and processing of visual information*. W.H. Freeman, San Fransisco.

Marslen-Wilson, W.D. 1987. Functional parallelism in spoken word recognition. *Cognition*, 25:71–102.

Mazuka, R. and K. Itoh. 1995. Can Japanese speakers be led down the garden path? In R. Mazuka and N. Nagai, editors, *Japanese Sentence Processing*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, chapter 13, pages 295–329.

Milward, D. 1994. Dynamic dependency grammar. *Linguistics and Philosophy*, 17:561–605.

Milward, D. 1995. Incremental interpretation of categorial grammar. In *Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics*, pages 119–126, University College, Dublin, March.

Mitchell, D.C., M.M.B. Corley, and A. Garnham. 1992. Effects of context in human sentence parsing: Evidence against a discourse-based proposal mechanism. *Journal of Experimental Psychology*, 18(1):69–88.

Moortgat, M. 1988. *Categorial Investigations*. Foris.

Partee, B., A. ter Meulen, and R.E. Wall. 1993. *Mathematical methods in Linguistics*. Dordrecht: Kluwer Academic Publishers.

Phillips, C. 1995. Right association in parsing and grammar. In C. Schütze, J. Ganger, and K. Broihier, editors, *Papers on Language Processing and Acquisition*, MIT Working Papers in Linguistics, volume 26. MIT, Cambridge, MA, pages 37–93.

Phillips, C. 1996. *Order and Structure*. Ph.D. thesis, Department of Linguistics and Philosophy, MIT, Cambridge, MA.

Pickering, M. and G. Barry. 1991. Sentence processing without empty categories. *Language and Cognitive Processes*, 6(3):229–259.

Pollard, C. and I. Sag. 1994. *Head-Driven Phrase Structure Grammar*. CSLI and University of Chicago Press, Stanford, Ca. and Chicago, Ill.

Pritchett, B.L. 1988. Garden path phenomena and the grammatical basis of language processing. *Language*, 64(3):539–576.

Pritchett, B.L. 1991. Head position and parsing ambiguity. *Journal of Psycholinguistic Research*, 20(3).

Pritchett, B.L. 1992. *Grammatical Competence and Parsing Performance*. University of Chicago Press, Chicago, IL.

Pulman, S.G. 1986. Grammars, parsers and memory limitations. *Language and Cognitive Processes*, 1(3):197–225.

Rayner, K., M. Carlson, and L. Frazier. 1983. The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of verbal learning and verbal behavior*, 22:358–374.

Schabes, Y., A. Abeillé, and A.K. Joshi. 1988. Parsing strategies with "lexicalized" grammars: Applications to tree adjoining grammars. In *12th International Conference in Computational Linguistics*, pages 578–583, Budapest, August.

Sells, P. 1994. Sub-phrasal syntax in Korean. *Language Research*, 30(2).

Sells, P. 1995. Korean and Japanese morphology from a lexical perspective. *Linguistic Inquiry*, 26(2):277–325.

Shieber, S. and M. Johnson. 1993. Variations on incremental interpretation. *Journal of Psycholinguistic Research*, 22(2):287–318.

Stabler, E.P. 1994a. Parsing for incremental interpretation. Manuscript, University of California at Los Angeles.

Stabler, E.P. 1994b. Syntactic preferences in parsing for incremental interpretation. Manuscript, University of California at Los Angeles.

Stevenson, S. 1993. A competition-based explanation of syntactic attachment preferences and garden path phenomena. In *Proceedings of the 31st Meeting of the Association for Computational Linguistics*, pages 266–273, Columbus, Ohio, June.

Stevenson, S. 1994a. Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research*, 23(4):295–321.

Stevenson, S. 1994b. A competitive attachment model for resolving syntactic ambiguities in natural language parsing. Technical Report 18, Rutgers, Center for Cognitive Science.

Sturt, P. 1994. Unconscious reanalysis in an incremental processing model. Manuscript, Centre for Cognitive Science, University of Edinburgh.

Sturt, P. and M. W. Crocker. 1995. Incrementality and monotonicity in syntactic parsing. In D. Milward and P. Sturt, editors, *Incremental Interpretation*, Edinburgh Working Papers in Cognitive Science, volume 11. Centre for Cognitive Science, The University of Edinburgh, chapter 2, pages 23–66.

Sturt, P. and M.W. Crocker. 1996. Monotonic syntactic processing: a cross-linguistic study of attachment and reanalysis. *Language and Cognitive Processes*, 11(5):449–494.

Sturt, P and M.W. Crocker. 1997. Thematic monotonicity.

Sturt, P. and M.W. Crocker. to appear. Generalized monotonicity for reanalysis models. to appear in J.D. Fodor and F. Ferreira, editors, *Reanalysis in Sentence Processing*, Kluwer Academic Publishers.

Thompson, H., M. Dixon, and J. Lamping. 1991. Compose-reduce parsing. In *Proceedings of the 29th Meeting of the Association for Computational Linguistics*, pages 87–97, Berkley, California, June.

Traxler, M.J. and M.J. Pickering. 1996. Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, 35:454–475.

Traxler, M.J., M.J. Pickering, and C. Clifton. 1996. Architectures and mechanisms that process prepositional phrases and relative clauses. Paper presented at the *Arcitectures and Mechanisms for Language Processing* conference, Torino, September.

Trueswell, J.C., M.K. Tanenhaus, and C. Kello. 1993. Verb-specific constraints on sentence processing: Separating effects of lexical preference from garden paths. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19(3):528–553.

Vijay-Shanker, K. 1992. Using descriptions of trees in a tree adjoining grammar. *Computational Linguistics*, 18(4):481–517.

Wall, R. 1972. *Introduction to mathematical linguistics*. Prentice Hall, Englewood Cliffs, New Jersey.

Warner, J. and A.L. Glass. 1987. Context and distance-to-disambiguation effects in ambiguity resolution: Evidence from grammaticality judgements of garden path sentences. *Journal of Memory and Language*, 26:714–738.

Weinberg, A. 1993. Parameters in the theory of sentence processing: Minimal committment theory goes East. *Journal of Psycholinguistic Research*, 22(3):339–364.

Weinberg, A. 1995. Licensing constraints and the theory of language processing. In R. Mazuka and N. Nagai, editors, *Japanese Sentence Processing*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, chapter 10, pages 235–255.

Yamashita, K. 1994. *Processing of Japanese and Korean*. Ph.D. thesis, Ohio State University, Columbus, Ohio.