EXTRACTION, MOVEMENT AND DEPENDENCY THEORY

Christian Kreps

Thesis submitted for the degree of Doctor of Philosophy
UNIVERSITY COLLEGE LONDON
Abstract

My aim in this thesis is to explore the potential of a dependency-based theory of syntax to account for extraction phenomena in English. Dependency grammars differ from constituency-based theories in expressing syntactic structure through direct relations between words rather than through their participation in phrase structure. However, whereas the majority of syntactic theories employing a phrase structure formalism have relatively well-developed accounts of extraction, dependency-based accounts have, in comparison, been virtually non-existent. This is surprising given that extraction phenomena constitute a key body of data, and an important test of a theory's capacity to account for long-distance syntactic relations. In order to account for these data theories have often suggested complex analyses involving multiple levels of representation and syntactic movement. However, in this thesis I hope to show that a monostratal theory of dependency with no such process of syntactic movement can provide a plausible and adequate account of extraction, which may compete on an equal footing with more complex approaches advanced in constituency-based theories.

After presenting a wide range of extraction data in the first chapter, in the second and third chapters I compare and contrast accounts of these data formulated within Principles and Parameters Theory, representing the constituency-oriented linguistic mainstream, and Word Grammar, the only dependency theory to offer any account of extraction phenomena. The fourth chapter will outline an alternative theory of dependency syntax, which I will refer to as Licensing Grammar (LG). Although LG takes Word Grammar as its starting point, it incorporates many significant differences from this and other theories, notably in its reformulation of dependency in terms of licensing. The fifth chapter then examines how LG, when integrated with a relatively simple parsing system, can serve as the basis of an effective account of extraction data.
Acknowledgements

This thesis does not just represent the outcome of my own endeavours over the last four years; it is also the product of the effort, dedication and tolerance of many other people. I would like to take this opportunity to thank them.

Firstly, my sincerest thanks go to Professor Richard Hudson. The title of ‘thesis supervisor’ hardly does justice to the influence Dick has had on my work, both through his own theory of Word Grammar and through his astute and perceptive responses to my own ideas. It was Dick who ‘converted’ me to dependency theory in the first place, and since then he has been an inspiring teacher who has guided my progress with patience, humanity and good humour. I will always be grateful for the time and effort he has spent dealing with my problems, answering my questions and reading just about everything that I have ever written. Perhaps one of the most difficult aspects of producing a Ph.D thesis is treading the sometimes narrow path between arrogance and humility; if I have managed to achieve the right balance here it is entirely due to Dick’s tolerance and wisdom.

I am also enormously indebted to Professor Neil Smith. It was Neil who first showed me what linguistics is really all about, and who taught me the value of parsimony and elegance, qualities which, I hope, have permeated this thesis to some extent. I am also grateful to the other members of the Linguistics department at UCL, particularly Rita Manzini and Hans van de Koot, who have been very helpful in commenting on my work and providing me with useful reading material. In addition I would like to thank the participants of the weekly Word Grammar seminar, notably Nik Gisbourne, Jasper Holmes, And Rosta and Dimitra Tzanidaki, who have all listened patiently to my rantings and have made valuable suggestions as to how some of the ideas therein might be improved (or discarded).

My studies at UCL would have been impossible without the four year postgraduate studentship I received from the Humanities Research Board of the British Academy. I gratefully acknowledge their support.

I would like to say a special thank you to my mother and father, as well as my sister and two grandmothers for helping me and supporting me in different ways, for putting up with me and for not asking (too often) when this work was going to be finished. I am particularly grateful for my father’s financial support throughout my studies. Many thanks too to my sister Melissa and Quentin Casares for their help with the hideous task of photocopying.
I am and will always remain especially grateful to my wife Dimitra who has also been a wonderful companion and friend. Through her loving care and practical help she has made the completion of this thesis much easier in many ways. Moreover, Dimitra’s completion of her own Ph.D in a mere two and a half years has stood as a constant reminder of what can be achieved if you put your mind to it.  

Ευχαριστώ Μικρή!  

It is no exaggeration to say that I would probably never have finished this thesis without the many great composers, both dead and alive, whose music has provided a joyful and inspiring soundtrack for much of my work. J.S Bach and Michael Nyman deserve special mention. Finally, a small thank you to my small cat Daisy who has kept me company throughout the long nights I have been writing.
Table of Contents

Abstract 2

Acknowledgements 3

Table of Contents 5

List of Abbreviations 9

1 - An Overview of Extraction 10

1.1 - Introduction 10

1.2 - The Data 11

1.2.1 - WH-Extraction 11

1.2.2 - Extraction within other contexts 14

1.2.3 - ‘Unbounded extraction’ 16

1.2.4 - Constraints on extraction - ‘islands’ 17

1.2.5 - Pragmatic constraints on extraction 19

1.2.6 - Extraction asymmetries 20

1.2.7 - Parasitic gaps 23

1.3 - Some Supplementary Data 24

1.3.1 - Island domains and tense 24

1.3.2 - More about the subject island 26

1.3.3 - Subjects versus objects 28

1.3.4 - Parasitic gaps revisited 32

1.4 - Extraction and Syntactic Theory 33

1.5 - Thesis outline 37
3 - Continued

3.4 - Word Grammar and Extraction Data

3.4.1 - Extraction and long-distance dependencies
3.4.2 - The visitor relation and propositions
3.4.3 - WH-word and head of clause
3.4.4 - Constraints on extraction
3.4.5 - Constraints on extractability

3.5 - Appraisal of the WG Account

4 - Licensing Grammar - Towards a New Theory of Dependency Syntax

4.1 - Introduction
4.2 - Dependency as Licensing

4.2.1 - The content of dependency
4.2.2 - Existential dependency and licensing
4.2.3 - Dependency without GRs

4.3 - Licensing Structure

4.3.1 - Some preliminaries
4.3.2 - Discovering licensing structure
4.3.3 - Finite and non-finite verbs
4.3.4 - LG and Categorial Grammar
4.3.5 - Words and elements
4.3.6 - Arguments and adjuncts

4.4 - Syntactic and Semantic Relations

4.4.1 - The mismatch between syntax and semantics
4.4.2 - Identifying Objects
4.4.3 - Subjects and d-command
4.4.4 - ‘Raising’ and ‘subject control’ structures
4.4.5 - ‘Object control’ and ‘ECM’ structures
4.4.6 - The syntax and semantics of adjuncts
4.4.7 - Word order in LG
4.4.8 - ‘Control’ phenomena in adjunct clauses

4.5 - Conclusions
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>176</td>
</tr>
<tr>
<td>5.2</td>
<td>Some Processing-Related Issues</td>
<td>177</td>
</tr>
<tr>
<td>5.2.1</td>
<td>The need for processing</td>
<td>177</td>
</tr>
<tr>
<td>5.2.2</td>
<td>A brief overview of parsing</td>
<td>179</td>
</tr>
<tr>
<td>5.2.3</td>
<td>Other processing-related accounts of extraction</td>
<td>183</td>
</tr>
<tr>
<td>5.3</td>
<td>LG and Parsing</td>
<td>185</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Some preliminary considerations</td>
<td>185</td>
</tr>
<tr>
<td>5.3.2</td>
<td>The structure of the processor</td>
<td>186</td>
</tr>
<tr>
<td>5.3.3</td>
<td>The operation of the LGP</td>
<td>191</td>
</tr>
<tr>
<td>5.3.4</td>
<td>Some advantages of pair-based parsing</td>
<td>196</td>
</tr>
<tr>
<td>5.3.5</td>
<td>Cell structure and reanalysis</td>
<td>199</td>
</tr>
<tr>
<td>5.4</td>
<td>Extraction and Valency Transfer</td>
<td>203</td>
</tr>
<tr>
<td>5.4.1</td>
<td>LG and the syntactic basis of extraction</td>
<td>203</td>
</tr>
<tr>
<td>5.4.2</td>
<td>Constraining VT</td>
<td>205</td>
</tr>
<tr>
<td>5.4.3</td>
<td>Local extraction and the Priority Principle</td>
<td>208</td>
</tr>
<tr>
<td>5.4.4</td>
<td>VT and long-distance extraction</td>
<td>212</td>
</tr>
<tr>
<td>5.4.5</td>
<td>The structure of WH-words</td>
<td>218</td>
</tr>
<tr>
<td>5.4.6</td>
<td>WH-extraction</td>
<td>223</td>
</tr>
<tr>
<td>5.5</td>
<td>Constraints on WH-Displacement</td>
<td>227</td>
</tr>
<tr>
<td>5.5.1</td>
<td>The nature of island constraints</td>
<td>227</td>
</tr>
<tr>
<td>5.5.2</td>
<td>The adjunct island and the co-ordinate structure island</td>
<td>228</td>
</tr>
<tr>
<td>5.5.3</td>
<td>The WH-island</td>
<td>234</td>
</tr>
<tr>
<td>5.5.4</td>
<td>The relative clause island and the complex NP island</td>
<td>238</td>
</tr>
<tr>
<td>5.5.5</td>
<td>Pre-dependent domains and the subject island</td>
<td>246</td>
</tr>
<tr>
<td>5.5.6</td>
<td>Summary of the island data</td>
<td>250</td>
</tr>
<tr>
<td>5.6</td>
<td>Extraction Asymmetries</td>
<td>252</td>
</tr>
<tr>
<td>5.6.1</td>
<td>Objects v adjuncts</td>
<td>252</td>
</tr>
<tr>
<td>5.6.2</td>
<td>Subjects v objects</td>
<td>263</td>
</tr>
<tr>
<td>5.6.3</td>
<td>A final note on parasitic gaps</td>
<td>268</td>
</tr>
<tr>
<td>5.7</td>
<td>Conclusions</td>
<td>271</td>
</tr>
</tbody>
</table>

References | 276  |
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNP island</td>
<td>Complex Noun Phrase island</td>
</tr>
<tr>
<td>CED</td>
<td>Constraint on Extraction Domains</td>
</tr>
<tr>
<td>CGF</td>
<td>Complement Gap Filter</td>
</tr>
<tr>
<td>DG</td>
<td>Dependency Grammar</td>
</tr>
<tr>
<td>ECP</td>
<td>Empty Category Principle</td>
</tr>
<tr>
<td>HPSG</td>
<td>Head-driven Phrase Structure Grammar</td>
</tr>
<tr>
<td>GPSG</td>
<td>Generalised Phrase Structure Grammar</td>
</tr>
<tr>
<td>GR</td>
<td>Grammatical Relation</td>
</tr>
<tr>
<td>LFG</td>
<td>Lexical Functional Grammar</td>
</tr>
<tr>
<td>LG</td>
<td>Licensing Grammar</td>
</tr>
<tr>
<td>LGP</td>
<td>Licensing Grammar Parser</td>
</tr>
<tr>
<td>PG</td>
<td>Parasitic Gap</td>
</tr>
<tr>
<td>PP Theory</td>
<td>Principles and Parameters Theory</td>
</tr>
<tr>
<td>PS Rule</td>
<td>Phrase Structure Rule</td>
</tr>
<tr>
<td>RC island</td>
<td>Relative Clause island</td>
</tr>
<tr>
<td>WG</td>
<td>Word Grammar</td>
</tr>
</tbody>
</table>
CHAPTER 1
AN OVERVIEW OF EXTRACTION

1.1 - Introduction

Of all the constructions of natural language, few have attracted as much attention as extraction. With the possible exception of passivisation, it is hard to think of any other linguistic phenomenon that has been so widely discussed within such a broad range of theoretical frameworks; within the modern generative tradition of linguistics, interest in extraction can be traced back to Chomsky (1964), who formulates an early version of the A-over-A Principle (Ross 1967) which, among other things, rules out extraction of a WH-element from a relative clause. Since then a significant amount of work has been carried out on extraction-related data within mainstream Chomskyan syntactic theory, and indeed it is no exaggeration to say that up until 1992, with the advent of the Minimalist Programme, these issues have in many ways been a defining influence on the course of the theory’s development from the Standard Theory through to the Principles and Parameters model. Extraction has, of course, also been widely discussed in and played an important part in the development of a number of other theories such as Generalised Phrase Structure Grammar (GPSG) (Gazdar 1981, Gazdar et al. 1985), Head-Driven Phrase Structure Grammar (HPSG) (Pollard and Sag 1987, 1994) and Lexical Functional Grammar (LFG) (Bresnan 1982).

However, all of the theories of grammar mentioned in the previous paragraph share one important characteristic; they are all constituency-based, in the sense that they seek to express syntactic structure in terms of configurational phrase patterns. However, in this thesis I will examine extraction from a different perspective, and I will instead explore how these data can be accounted for in a Dependency Grammar. Dependency Grammar, or DG, is a generic term which describes any theory of language which expresses syntactic knowledge in terms of direct relationships between words rather than through phrase structure. DG in its purest form is represented by only three contemporary theories of syntax, Word Grammar (Hudson 1984-a, 1990), Lexicase (Starosta 1988) and Mel’čuk’s (1988) theory. However, various other frameworks also make use to some extent of a dependency-related formalism (Hudson 1993). This is true, for example, of Case Grammar (Fillmore 1968, Anderson 1971, 1977), Relational Grammar (Perlmutter 1983) and Categorial Grammar (Oehrle et al. 1988, Wood 1993).
It is fair to say that extraction phenomena have received very little attention in DG compared to constituency-based theories. This is one of the reasons for which, in my opinion, no entirely adequate dependency-based account of these data has ever been put forward. This is particularly surprising given that a theory's treatment of the long-distance syntactic relations arising from extraction is often considered to be a key benchmark of its adequacy. My central concern in this thesis, then, is to demonstrate that a model of dependency syntax, when supplemented by processing considerations, is indeed capable of offering a plausible and satisfactory account of extraction phenomena, and one which compares favourably with the treatment of these data in other theories. As far as I am aware, this is the first detailed evaluation of a DG's capacity to account specifically for extraction data.

In sections 1.2 and 1.3 of this chapter I will present a full range of relevant English data, including constraints on extraction such as 'island' phenomena. Section 1.4 will then discuss some of the implications of these data for syntactic theory, while also presenting a very brief outline of dependency and constituency-based syntax.

1.2 - The Data

1.2.1 - WH-extraction

English, like many other languages, displays a phenomenon commonly known as WH-movement or WH-extraction. This is a syntactic construction which involves the occurrence of a so-called 'WH-word', such as "who", "what", "why" or "how" in a clause-initial position in certain environments, notably matrix interrogative structures such as those in (1):

1)  i. What did Balanescu play?
    ii. Why did Sid eat a beetle sandwich?

These WH-words are said to be 'moved' or 'extracted' because they occur in a clause-initial position which they would not otherwise occupy. In (1i), for example, "what" is the object of the verb "play"; objects in English generally occur immediately after their governing verb as in (2), a non-WH equivalent of (1i):

2)  Balanescu played 'Angelfish Decay'.

In some intuitive sense, then, a fronted WH-word could be said to belong elsewhere in its sentence. One way of expressing this intuition is to allow each displaced WH-word to be
associated with a gap in the subsequent structure, the gap occurring where the WH-word would appear under normal circumstances. Thus in the example in (1i) we could say that a gap occurs immediately after the verb “play”, a space usually occupied by the verb’s object, such as “Angelfish Decay” in (2) above. This gap is illustrated in (3):

3) What did Balanescu play _?

The terms “movement” and “extraction” are somewhat controversial; it isn’t necessarily the case that WH-words actually move to their clause-initial position. The gap in (3) might serve only to indicate where an element would ‘normally’ occur. To say that a word has moved implies that at some point it actually occupied the position signified by the gap. Thus terms such as ‘movement’ and ‘extraction’ imply a derivational process whereby a WH-word vacates its original position and moves to another location. This, in turn, raises a number of issues, not least of which is the question concerning the point at which a ‘moved’ word occupies its original position. The implication is that each observable sentence is associated with at least one underlying, abstract structure in which certain words may occupy different positions from their surface order. It is thus no accident that the terms ‘movement’ and ‘extraction’ have generally emanated from theories where such notions of abstract, underlying linguistic structure are indeed prevalent. In this sense, then, they could be classed as theory-internal terms, the use of which implicitly subscribes to a specific approach to language. If, however, I continue to refer to ‘movement’ and ‘extraction’, it is purely for reasons of convenience, and in no way implies that I favour a “movement analysis” of these constructions.

The displacement of questioned WH-words in English is not compulsory, and it is possible for a WH-word to occupy what would be its canonical position in the clause:

4) i. Balanescu played what?
   ii. Who played what?

(4i) is a marked option in English, and could only be uttered within a specific discourse context, expressing surprise or incredulity, and with a heavy intonational emphasis placed on “what”. In (4ii), however these effects are absent; English, in common with most other languages displaying WH-extraction, allows only one WH-element to be displaced per clause. In cases where two elements are questioned, as in (4ii), then evidently one of these will have to remain in situ, though without attracting the emphasis and echoic interpretation associated with “what” in (4i). In spite of examples like (4), then, the point stands, that English is a
language in which WH-elements tend to ‘move’ to a clause-initial position. Note, though, that WH-displacement in English, as in all other languages, is uniquely leftward. That is to say that no language has yet been attested where WH-words are systematically displaced to the end or to the middle of a sentence. It would thus appear that a displaced WH-word will always precede its associated gap (Chomsky 1973).

As already illustrated in (1) above, WH-extraction affects both arguments and adjuncts in a similar way. Of course the position of adjuncts in a sentence tends to be rather more free than that of arguments, and it is not always necessarily easy to associate a ‘fronted’ WH-adjunct with any particular gap in the subsequent sentence:

5) i. Why did Sid eat a beetle sandwich?
   ii. Sid ate a beetle sandwich in order to impress Sally.
   iii. In order to impress Sally, Sid ate a beetle sandwich.

Nevertheless, it is clear that WH-adjuncts are displaced in the same sense as their argument counterparts. Indeed, unlike arguments, it is actually quite difficult to find examples of WH-adjuncts in situ; compare (6) below with (4) above:

6) i. *Sid ate a beetle sandwich why?
   ii. /*They got him to hospital how?
   iii. 0*He was operated on when?

Even with considerable emphasis placed on the in situ WH-adjuncts here the sentences seem at best slightly unnatural (6iii), or worse in the case of (6i) and (ii). These sentences thus provide us with an early example of a contrast between WH-arguments and WH-adjuncts which will be a recurring theme throughout this thesis.

Arguments and adjuncts of verbs are not the only constituents which can be displaced in English. The argument of a preposition, as well as the entire prepositional phrase itself, can also be extracted:

---

1In languages lacking extraction WH-words always remain in situ, Japanese and Turkish being two well-known examples.

2In accordance with common practice I will use a series of prefixed symbols to indicate the degree of ungrammaticality associated with a sentence. In order of ascending ill-formedness these symbols are: ?, ?/*, *, and **.
7) i. Which sack did Santa hide the cash in _ ?
   ii. In which sack did Santa hide the cash _ ?

In (7i) the fronted WH-phrase “which sack” functions as the complement of the preposition “in” whereas in (7ii) the entire WH-phrase “in which sack” constitutes a displaced adjunct of the verb “hide”.

1.2.2 - Extraction within other contexts.

Extraction is not confined to matrix interrogative clauses; a similar phenomenon also occurs in Relative Clauses (8) and in certain complement clauses, sometimes known (rather confusingly) as indirect questions (9):

8) i. They loved the piece [which Balanescu played _ ].
   ii. I washed the sack [in which Santa hid the cash _ ].

9) i. I wonder [what Balanescu played _ ].
   ii. He asked [where Santa had hidden the cash _ ].
   iii. I know [why Sid ate the beetle sandwich _ ].
   iv. We couldn’t decide [how to get him to hospital _ ].

Clearly, the examples in (8) and (9) parallel the matrix interrogatives discussed earlier, in that they each involve a similar displacement of a WH-element to the clause-initial position and a resulting gap in the subsequent clause. Consequently, although the main emphasis of this thesis is on displacement within matrix WH-interrogatives, much of what I will go on to say will also be of direct relevance to other displacement contexts as well, such as indirect questions and relative clauses.

Paradoxically, perhaps, WH-words are not the only elements that may undergo WH-style displacement; the so-called topicalised elements underlined (10) below are all displaced in a similar way:

10) i. The chisel we’ll certainly need _ .
    ii. Sid I like _ . (not Ted)

---

3 Strictly speaking, the general term ‘indirect question’ is a bit misleading; often these embedded structures have nothing to do with interrogatives, as can be appreciated by the range of possible matrix verbs selecting this kind of WH-clause complement, including “realise” and “know” and “decide”.

14
It is by no means clear that these extracted elements are always topicalised; indeed “Sid” in (10ii) is more likely to be interpreted as a focus of the sentence. Nevertheless the displacement of these words is exactly parallel to that of WH-words, and, once again, although I will concentrate on displacement within matrix interrogative clauses, much of what I will say will also be directly relevant to ‘topic’ constructions.

There is, however, one important difference between matrix interrogatives and the other extraction contexts mentioned above; only the former trigger the inversion of the following subject with a tensed auxiliary verb. If an auxiliary verb is present in the sentence, then this will be inverted with the subject (11). If, however, no auxiliary would otherwise occur, then “do” is inserted (12).

11) i. The spectators were vandalising their seats.
   ii. What were the spectators vandalising?
   iii. *What the spectators were vandalising?

12) i. Sid ate a beetle sandwich.
   ii. Why did Sid eat a beetle sandwich?
   iii. *Why Sid ate a beetle sandwich?

In embedded WH-constructions, however, as well as in examples of ‘topic’ extraction, displacement does not trigger the inversion of an auxiliary verb and the subject, and the order of elements subsequent to the extracted WH-word is as expected:

13) i. I met the man whom Paddy was bribing.
    ii. *I met the man whom was Paddy bribing.

14) i. I don’t understand why Sid ate a beetle sandwich.
    ii. *I don’t understand why did Sid eat a beetle sandwich.

15) i. The chisel we will certainly need.
    ii. *The chisel will we certainly need.

The picture is further complicated by the fact that even in matrix interrogatives WH-subjects do not trigger inversion, hence the contrast between (16i) and (ii):

16) i. Who wrote ‘Angelfish Decay’?
    ii. *Who did write ‘Angelfish Decay’?

\(^4\) (16ii) is ungrammatical under the intended interpretation, without emphasis on the auxiliary verb.
Clearly then the phenomenon of 'auxiliary-inversion', and its occurrence in some contexts but not others, represents another facet of extraction phenomena which will have to be explored and accounted for within any serious treatment of these data.

1.2.3 - 'Unbounded extraction'

All the examples of extraction examined so far have involved the presence of a WH-word at the beginning of the clause and the subsequent occurrence of an associated gap in the same clause. In this sense, then, extraction could be said to be 'clause-bound', in that no WH-word has yet been seen to occur in a clause in which it doesn’t bear some relation to another word, either as an argument or adjunct. However, English differs from some languages in that it does actually allow 'long-distance extraction' where a WH-word occurring at the beginning of one clause may be associated with a gap in another clause. As an example of this consider (17) where “whom” occurs at the beginning of the matrix clause, even though it clearly functions as the object of ‘bribed’, the verb in the complement clause:

17) Whom did Paddy say [that he had bribed _ ]?

Described in movement-type terminology, “who” could be said to have moved out of the complement clause, enclosed in square brackets, and up into the matrix clause. This displacement thus crosses a clause boundary. Indeed, the distance between a fronted WH-word and its associated gap may potentially be far greater:

18) What does Balanescu suspect [that Wally believes [that Sid had told us [that Nyman found _ on his doorstep]]] ?

The example in (18) is grammatical, if cumbersome, even though the distance between the extractee “what” and its gap spans three clause boundaries. WH-adjuncts can also undergo long-distance displacement:

19) When did Nancy say [that she would visit Sid in hospital _ ]?

Examples such as (17) and (19) are commonplace, and it is for this reason that WH-extraction is sometimes referred to as an example of 'unbounded movement', an operation which contrasts with most other grammatical processes and relationships, which are generally subject to strict locality restrictions. This apparent unboundedness of extraction applies equally to all displacement contexts:
20) i. I wonder [who he said [that they thought [Paddy had bribed _]]].
   ii. I found the place [where they said [that Santa had hidden the cash _]].
   iii. Nyman I know [the critics despise _].

The long-distance nature of this type of displacement is one of the most unusual aspects of the phenomenon, and one of the reasons why extraction data pose a unique challenge for any theory of syntax.

1.2.4 - Constraints on extraction - 'islands'

In spite of the apparently unbounded nature of extraction, there are in fact a number of important factors which serve to constrain its application. Indeed, for almost as long as there has been serious discussion of displacement phenomena a key component of this discussion has been concerned with the factors restricting both the class of elements that may be displaced and the range of domains from which displacement may take place. For example, in what is, I think, one of the first references to constraints of this sort Chomsky (1964) discusses the ungrammaticality arising from the extraction of a WH-word from a relative clause. This is illustrated in (21):

21) i. I met a woman [who knew Einstein].
   ii. *Who did you meet a woman [who knew _]?  

Ross (1967) is the first systematic survey of constraints obtaining on extraction. He identifies a series of 'islands', domains from which displacement is prohibited; these include 'complex Noun Phrases' (complement clauses of nouns), subject domains and co-ordinate structures, amongst others. Examples of these 'islands' are given below:

22) i. I've heard [a rumour that Sid was taken to hospital].
   ii. *Where have you heard [a rumour that Sid was taken _]? (complex NP island)
23) i. [Chasing sheep] tires Mr. Neville.
   ii. *What does [chasing _] tire Mr. Neville? (subject island)
24) i. Balanescu played 'Angelfish Decay' [and the neighbours rang the police].
   ii. *Who did Balanescu play 'Angelfish Decay' [and the neighbours rang _]? (co-ordinate structure island)

Notice, though, in connection with the co-ordinate structure island, that extraction becomes acceptable when applied to both conjuncts simultaneously:
25) i. Polly learned the 'Bird List Song' [and sang it in her shed].
   ii. What did Polly learn [and sing _ in her shed]?

This is sometimes known as 'Across the Board' (ATB) extraction.

The constraints on extraction illustrated in (21)-(24) apply to all displacement contexts, Relative Clauses (26i), 'indirect questions' (26ii) and 'topic' constructions (26iii):

26) i. "They found the cash which I've heard [a rumour that Santa stole _].
    ii. *I know who Balanescu played 'Angelfish Decay' [and the neighbours rang _]
    iii. *Sheep [chasing _] tires Mr. Neville.

Clearly, then, certain structures seem to attenuate severely the apparent unboundedness of extraction, and it is evidently not the case that the relationship between an extractee and its associated gap may span any clause boundary.

In addition to these well-known examples first discussed by Ross (1967), subsequent work has unearthed another range of properties and structures which constrain extraction. For example, displacement is not allowed from a clause which itself contains a fronted WH-word:

27) i. I know [who stole the cash].
    ii. *What do you know [who stole _]?
    iii. They asked [how Pulat Fallari killed his brother].
    iv. *Who did they ask [how Pulat Fallari killed _]?

This constraint has come to be known as the 'WH-island' (Chomsky 1981, Rizzi 1982). Most adjunct domains also constitute an island from which displacement is impossible:

28) i. Sid ate a beetle sandwich [because he wanted to impress Sally].
    ii. *Whom did Sid eat a beetle sandwich [because he wanted to impress _]?

29) i. The neighbours called the police [after Balanescu played 'Angelfish Decay'].
   ii. *What did the neighbours call the police [after Balanescu played _]?

However, extraction from an adjunct is possible in certain circumstances, notably when the adjunct itself is a simple prepositional phrase:

30) i. Nigel threw up [in his euphonium].
    ii. What did Nigel throw up [in _]?
1.2.5 - Pragmatic constraints on extraction

In addition to the structurally-defined 'island' domains, such as subject, adjunct and complex NP discussed above, other, apparently more pragmatic, factors are also known to have an effect on extraction. These factors include negation (Ross 1984) and factivity (Cinque 1991); the complement clause of a negated or a factive verb is said to constitute an island, although many of these cases are by no means clear-cut:

31)  i. I don't think [Santa hid the cash in his sock].
    ii. What don't you think [Santa hid _ in his sack]?

32)  i. I regret [that I spent Christmas in Albania].
    ii. Where do you regret [that you spent Christmas _ ]?

So too it has been argued (Manzini 1992) that definiteness also has an effect on displacement from within nominals:

33)  i. I bought [a/the portrait of Colonel Gadaffi].
    ii. Who did you buy [a portrait of _ ]?
    iii. *Who did you buy [the portrait of _ ]?

As Ross (1967) notes, extraction of the complement of "of" in (33ii) and (33iii) is predicted to violate the Complex NP Island. However it is generally felt that this really only applies to definite NPs (33iii) rather than their corresponding indefinites.

The status of these factive, negative and definite domains as true islands is debatable, and for this reason they are sometimes known as 'weak islands' (Manzini 1994-c). Neither (31ii) nor (32ii) seems particularly bad, especially when set within a suitable context, and it isn't difficult to think of perfectly grammatical analogues of (33iii):

34) Which country did he visit [the capital of _ ]?

It would appear then that any oddness associated with (31ii), (32ii) and (33iii) is not really structural or syntactic in nature, but arises instead from pragmatic considerations which are, at least in part, dependent on context. However, to describe these effects as pragmatic is not to dismiss them; it may be that these influences on extraction are elusive and ill-defined but they exist nevertheless. Moreover, pragmatic effects such as these, although they may not fall within the domain of syntax proper, clearly do interact closely and in interesting ways with syntactic phenomena. For example Lakoff (1986), Kuno (1987) and Deane (1989, 1991)
describe how certain island effects can be ameliorated or even neutralised altogether by pragmatic factors:

35) i. *How much money did they discuss [the claim that Santa stole _ ]?
    ii. *How much money did they make [the claim that Santa stole _ ]?

In addition, Erteschik-Schir and Lappin (1979) identify a series of verbs which seem to prevent extraction out of a normal complement clause, which would not otherwise be predicted to be an island:

36) *What did Ted whisper that he had bought _?

Once again, the pragmatic nature of the effect in (36) is underlined by the fact that within the right context it can become more acceptable:

37) (Ted was bragging to everyone that he had bought an Aston Martin, but whispered to us afterwards that he had really bought a moped).

What did Ted whisper that he bought _?

Traditionally, syntactic theories have abstracted away from pragmatic effects such as these and opted instead to present an ideal account of the 'core' syntactic cases (Chomsky 1986-a, Deane 1991, Kluender 1992). This is not necessarily an unreasonable position; there is, after all, no reason why a theory of syntax should have to account for pragmatic phenomena. Nevertheless some of the extraction data described above do attest to an interplay between syntax and pragmatics, and it is quite possible that the relationship between the two may be closer than some may care to admit. For this reason, ideally, a syntactic theory might seek at least to provide an interface with which a principled theory of pragmatics could be linked. In this way it might be possible, if nothing else, to keep alive the hope that some sort of holistic account for all aspects of extraction might one day be offered.

1.2.6 - Extraction asymmetries

Apart from the various restrictions on the domains from which extraction may take place, a great deal of work (following Huang 1982) has also concentrated on apparent asymmetries concerning the nature of the extractee itself. For example, when displaced from an island domain, adjuncts tend to give rise to a significantly worse degree of ungrammaticality than arguments do. Compare (38ii)/(39ii) with (38iii)/(39iii):
38) i. I heard a rumour [that Pulat Fallari killed his brother with a pencil].

ii. *Who did you hear a rumour [that Pulat killed _ with a pencil]? 

iii. **How did you hear a rumour [that Pulat killed his brother _]? 

39) i. I know [when Santa stole the cash].

ii. *What do you know [when Santa stole _]? 

iii. **When do you know [what Santa stole _]? 

Indeed, the so-called 'weak islands' discussed in the previous section often allow the perfectly grammatical extraction of an argument but not an adjunct. Thus (40iii), unlike (40ii) is ungrammatical: 

40) i. I don't believe [that Santa stole the cash in order to buy an Uzi].

ii. What don't you believe [that Santa stole _ in order to buy an Uzi]? 

iii. *(*)Why don't you believe [that Santa stole the cash _]? 

Here once again we see another example of a general asymmetry between adjuncts and arguments with respect to displacement; while adjuncts, unlike most arguments, generally cannot be extracted from, so too adjuncts tend to be less 'extractable' than arguments. In fact the situation is more complex than this, and in English the displacement of subjects from island domains is also generally significantly worse than the extraction of a complement from a similar domain; whereas the examples in (41ii) and (42ii) below are bad, those in (41iii) and (42iii) are totally uninterpretable: 

41) i. I know [where Santa hid the cash].

   ii. *What do you know [where Santa hid _]? 

   iii. **Who do you know [where _ hid the cash]? 

42) i. The neighbours rang the police [after Balanescu played 'Angelfish Decay'].

   ii. *What did the neighbours ring the police [after Balanescu played _]? 

   iii. **Who did the neighbours ring the police [after _ played 'Angelfish Decay']? 

In this way subjects would appear to pattern with adjuncts with respect to their extractability, at least in certain circumstances. As far as 'weak islands' are concerned, however, subjects behave more like complements in that they too can be displaced from these domains without any real difficulty:
43) i. I don’t believe [Santa stole the cash].  
   ii. What don’t you believe [Santa stole _ ]?  
   iii. Who don’t you believe [ _ stole the cash]?  

There is, in addition, one particular constraint on extraction which only affects subjects, 
the so-called ‘that-trace filter’ (Chomsky and Lasnik 1977). This prohibits the displacement 
of a subject which occurs immediately after the complementiser “that”. In other words, the 
constraint rules out structures where a subject gap occurs immediately after the 
complementiser (44ii). Generally, though, the extraction of a subject from an embedded clause 
is otherwise unproblematic (44i):  

44) i. Who do you think [ _ stole the cash]?  
   ii. *Who do you think [that _ stole the cash]?  

The ‘that-trace filter’ has no effect on the extraction either of complements or adjuncts:  

45) i. What do you think [that Santa stole _ ]?  
   ii. When do you think [that Santa stole the cash _ ]?  

Note, however, that the presence of the complementiser is only relevant to the displacement 
of an immediately adjacent subject; a subject embedded further down in a structure can be 
displaced across a complementiser without any problem:  

46) Who did Wally tell us that he thought [ _ had eaten a beetle sandwich]?  

In this way a complex pattern emerges concerning the ‘extractability’ of different 
elements according to the grammatical relation they bear in their clause; generally, subjects and 
adjuncts seem to pattern together, their displacement tending to be more problematic than that 
of complements. Occasionally, though, as in the case of ‘weak islands’, subjects seem to 
behave more like complements in being more extractable than adjuncts. Sometimes 
complements and adjuncts behave alike, as regards their immunity to ‘that-trace’ effects which 
affect subjects, for example. In fact, as I will go on to argue later in section 1.3.3, there is good 
reason to think that subjects and objects are actually more similar than they might initially 
appear, and that all arguments, unlike adjuncts, basically behave alike with respect to their 
extractability.
One phenomenon that has received a great deal of attention within a wide range of theories is the so-called 'parasitic gap', or PG (Chomsky 1981, 1982, Engdahl 1983, Hudson 1984-b, Manzini 1992, 1994-a). As I have already pointed out, extraction generally involves the association of a fronted WH-word with a 'gap' in the subsequent sentence, the gap informally signifying the place in which the WH-word would otherwise occur had it not been extracted. However, under certain circumstances it is possible for a displaced WH-word to be associated with two gaps in the subsequent structure:

47)  
   i. What did Wally buy _ in order to wear _ at the party?
   ii. Which drugs did Santa buy _ before trying _ first?

In each of the above structures the fronted WH-word seems to serve as the object of two verbs. This apparent multiple connection between an object and two verbs is usually thought to be restricted to WH-contexts. Generally non-WH equivalents of (47) are not possible:

48)  
   i. *Wally bought a fez in order to wear _ at the party.
   ii. Wally bought a fez in order to wear it at the party.

49)  
   i. *Santa bought the cocaine before testing _ first.
   ii. Santa bought the cocaine before testing it first.

For this reason, some analyses take the second gap in examples like (47) to be parasitic on the first gap. That is, the gap with which a displaced WH-word must be associated can itself license the presence of another, 'parasitic' gap. If, however, there is no initial gap, as in (48) or (49) for example, then no second 'parasitic' gap can be licensed (though see Hudson 1984-b for an alternative view).  

Parasitic Gaps have been problematic for most theories of grammar, and people have had to resort to often very complex syntactic machinery in order to account for these constructions. Whatever one’s approach to PG’s, however, they clearly lie within the realm of extraction data, and should ideally be treated in a unified way with these broader issues.

[^5]: There is, however, a single exception to this rule, a unique and problematic example which appears to involve the dependence of a single, non-WH object on more than one verb simultaneously:

   “Wally bought a fez to wear _ at the party”
1.3 - Some Supplementary Data

In the previous section I presented a reasonably comprehensive overview of extraction data in English. However, even from this relatively brief review it should be clear that facts relating to displacement and its associated constraints are often ill-defined and poorly understood. For example, one of the key characteristics of extraction phenomena is the almost infinite scalability of grammaticality judgements associated with violations of the various island constraints. Moreover, there remain wide areas in the realm of displacement and its related restrictions which so far lack any principled explanation in any syntactic theory. It should thus come as no great surprise that even now, following some 30 years of detailed research into these issues, it is possible to unearth new properties and factors which have a direct bearing on extraction, but which, as far as I know, have not yet been discussed within any framework. In this section I will present a small selection of these ‘undiscovered’ factors which are, I believe, of sufficient importance to deserve an explanation in any serious account of extraction.

1.3.1 - Island domains and tense

It has been known for some time that tense/finiteness\(^4\) has an effect on the status of some island domains. For example, Chomsky (1986-a) cites examples of the WH-island whose capacity to block extraction seems to be partially dependent on whether or not it is tensed.

50) i. *What did he tell you [when Sid ate _]? 
ii. ?What did he tell you [when to eat _]? 

It seems to me, however, that the true significance of tense as a factor in displacement and ‘islandhood’ is an issue that has been underestimated. While Chomsky (1986-a) and Manzini (1992, 1994-c) discuss the effects of tense on the WH-island, it seems that the absence of tense in any domain effectively annuls its status as an island. The bracketed, non-tensed domains in (51)-(53) should all constitute islands, and yet extraction from these domains seems to be almost perfectly acceptable:

51) i. Wally kept [his promise to wear a fez]. (Complex NP) 
ii. ?What did Wally keep [his promise to wear _]?

\(^4\)I will generally use the term ‘tense’ to refer to finiteness as well.
52) i. Sandra knows [how to play ‘Swan Rot’ on the violin]. (WH-clause)
   ii. Which piece does Sandra know [how to play _ on the violin]?

53) i. Pulat used a pencil [to kill his brother]. (Adjunct)
   ii. ?Whom did Pulat use a pencil [to kill _]?

The tensed equivalents of (51-53), however, shown in (54) below, clearly do function as islands to WH-displacement:

54) i. ?*What did wally keep [his promise that he would wear _ ]?
   ii. ?*Which piece do you know [how Sandra played _ ]?
   iii. *Whom did Pulat use a pencil [so he could kill _ ]?

In addition, at least some of the apparently pragmatically-oriented exceptions to the co-ordinate structure island discussed by Lakoff (1986) and Deane (1991) can also be accounted for by this same generalisation; extraction from the non-tensed conjunct in (55ii) is more acceptable than from its tensed equivalent in (56ii):

55) i. He went to the shop [and bought a fez].
   ii. What did he go to the shop [and buy _ ]?

56) i. He went the shop [and he bought a fez].
   ii. *What did he go to the shop [and he bought _ ]?

Note, however, that the increased ungrammaticality of (56ii) will arise, at least in part, from the fact that the first conjunct of a co-ordinate structure cannot be questioned in isolation:

57) *Did he go to the shop [and he bought a fez]?

Of course, we cannot comment on the extractability of subjects from non-tensed islands owing to the absence of subjects in non-finite environments. However, it should be pointed out that extraction of adjuncts from non-tensed islands generally appears to be equally problematic as extraction from corresponding tensed domains:

58) i) *When did Wally keep [his promise to wear a fez _ ]?
   ii) *Where did Pulat use a pencil [to kill his brother _ ]?

Although even here it is possible to find exceptions:

59) How does Brenda have [a duty to serve her country _ ]?
Clearly, then, there are grounds for believing that the role of tense as a factor in ‘islandhood’ needs to be re-examined.

1.3.2 - More about the subject island

Another aspect of displacement which, in my opinion, deserves more attention is the subject island. The properties of subject domains seem to be substantially different from those of other islands. For one thing, subject clauses seem to constitute the only island domain where the tense/non-tense differentiation discussed above does not appear to play any significant role; extraction from a non-tensed subject island, as in (60) and (61) below, results in ungrammaticality, unlike displacement from the various non-tensed island domains illustrated in (51)-(53) and (55) above:

60) i. [Chasing sheep] tires Mr. Neville.
   ii. *What does [chasing _] tire Mr. Neville?

61) i. [To meet Elvis] would make Noam happy.
   ii. *Who would [to meet _] make Noam happy?

More importantly, perhaps, it is worth examining why exactly a subject domain should constitute an island in the first place; we know that complement clauses in English routinely allow extraction, so why should subject clauses be any different? Is there some inherent property of subjects which makes them islands to displacement, or is there instead some incidental fact concerning subjects which happens to be responsible for the effect? As I will go on to show in the next chapter, standard analyses within Principles and Parameters Theory have tended to concentrate on the former approach, linking the properties of subjects to their fundamental configurational definition. However, there seems to be some evidence favouring the second approach; more specifically, the fact that subjects are islands in English would appear to be closely linked with their canonical position in a sentence, before the matrix verb. In general there seems to be an overriding constraint preventing extraction from any subordinate clause which precedes the matrix verb but allowing extraction from any clause.

---

7 Displacement from a tensed subject clause gives rise to horrendous effects, though this is partly due to a general ban on embedded tensed subject clauses (Hawkins 1994):

   i. [That Noam met Elvis] seems highly improbable to me.
   ii. **Who does [that Noam met _] seem highly improbable to you?
   iii. *Does [that Noam met Elvis] seem improbable to you?
which follows it. Thus, for example, in English, when a subject clause is extraposed - placed at the end of the matrix clause - displacement from it becomes acceptable:

62) i. What does it tire Mr. Neville [to chase _]?
   ii. Who would it make Noam happy [to meet _]?

   Similar evidence comes from languages where no extraction can take place from any subordinate clause which precedes the matrix verb, irrespective of whether the clause in question functions as a subject or as an object. Thus in Modern Greek*, for example, extraction from both a subject and object clause is fine when they follow the matrix verb:

63) i) Ποιόν φημολογείται [δότι ο Κουλής σκότωσε];
   Whom is rumoured [that Coulis killed _]sub?
   Whom is it rumoured that Coulis killed?

   ii) Ποιόν πιστεύεις [δότι ο Κουλής σκότωσε];
   Whom you believe [that Coulis killed _]obj?
   Whom do you believe that Coulis killed?

However, extraction from the same subject and object clauses becomes ungrammatical when they precede the matrix verb:

64) i) *Ποιόν [δότι ο Κουλής σκότωσε] φημολογείται;
   Whom [that Coulis killed _]sub is rumoured?

   ii) *Ποιόν [δότι ο Κουλής σκότωσε] πιστεύεις;
   Whom [that Coulis killed _]obj you believe?

In both (63i) and (64i) the subordinate clause functions as the subject of the main verb "φημολογούμαι" - 'to be rumoured'. In (63ii) and (64ii) the same clause functions as the object of the verb "πιστεύω" - 'believe'. The data in (63) and (64) would thus appear to lend strong support to the claim that it is the position of a domain in relation to the verb which is responsible for determining 'islandhood' rather than any inherent property. This generalisation would appear to apply equally to adjunct clauses as well as argument domains; contrast (65i) below with (65ii):

---

8In Modern Greek, a ‘free word-order’ language, subjects and objects, whether clausal or not, may occur anywhere in relation to the verb (Joseph and Philippaki-Warburton 1987, Tzanidaki 1995, 1996).
65)  i) *Whom did King Zog leave early [in order to meet _ ]?  
ii) *Whom [in order to meet _ ] did King Zog leave early?

As we have already seen in (53ii) above, extraction of a complement from a non-tensed adjunct domain as in (65i) is relatively unproblematic. However, when the same adjunct is placed before the matrix verb then extraction is forbidden.

It would seem, then, that there may be some overriding principle which prevents extraction out of any domain which precedes its head. The generalisation predicts the ungrammaticality of extraction from subject clauses in English - at least when the subject occurs in its canonical position before the matrix verb. However, if this generalisation holds on a wider basis, as the evidence from Greek would seem to suggest, then it might also serve to account for the fact that, as far as I know, extraction does not occur in SOV languages such as Turkish and Japanese. Evidently in these languages both subject and object domains will generally precede the matrix verb - thus rendering them islands according to the generalisation formulated above. Conversely, VSO languages are predicated routinely to allow displacement from both subject and object clauses. This predication is, in fact, borne out, at least in the case of Chamorro (Chung 1983).

Of course, the generalisation that extraction is forbidden from pre-head domains remains just that, a generalisation. Any language which permitted displacement from subordinate clauses which precede their head would serve to disprove any claim the generalisation had to universality. According to Tomlin (1986) just over half of the world’s languages are SOV. It could well be that amongst this set there is just such a language that does indeed violate this generalisation. More importantly, even if this generalisation can be shown to hold (semi-) universally, what still remains to be done is to explain why it holds in the first place. I will say more about this issue in chapters 3 and 5.

1.3.3 - Subjects versus objects

Recall in section 1.2.6 I discussed a notable disparity between arguments and adjuncts with respect to their extractability; when displaced from island domains adjuncts were shown consistently to give rise to a more severe degree of ungrammaticality than complements, as illustrated in examples (38) - (40) above. However, a similar asymmetry was also shown to exist between subjects and objects; extraction of a subject from most (but not all) islands results in worse effects than extraction of an object from the same domain:
For this reason subjects have often been classed together with adjuncts as far as their 'extractability' is concerned (Chomsky 1981, 1986-a, Lasnik and Saito 1984, 1990, Rizzi 1990, Manzini 1992).

There are, however, a number of problems with this approach. Firstly, apart from extraction data, subjects would otherwise appear to have nothing in common with adjuncts, and thus to bring the two together as some sort of natural class is perhaps rather an expensive way of dealing with a limited range of data. Secondly, the apparent extraction asymmetries between subjects and objects only apply in certain circumstances; as I noted in section 1.2.6, unlike adjuncts, all arguments (subjects and objects) can be extracted equally well from the so-called 'weak islands' (Manzini (1994-c):

Moreover, the effects of the 'that-trace filter', discussed in section 1.2.6, apply solely to subjects, setting them apart both from adjuncts and complements. Thus the ungrammaticality of (70ii) below arises from the fact that the subject 'gap' occurs immediately after the complementiser “that”:

In my opinion, the apparent complexity of this situation is really the result of the data being viewed in the wrong way. I believe that what appears to be an inherent asymmetry between subjects and objects with respect to their extractability is only apparent, and can
actually be explained by extending the 'that-trace filter' to something analogous to a 'complementiser-gap filter'. In this way the complex set of data outlined above can be shown to be the product of an interaction between just two more simple constraints, a pervasive contrast between adjuncts and arguments, and an incidental constraint barring the extraction of any subject which is adjacent to a complementiser.

Consider first of all the violations of the complex NP island in (71) below:

71)  
   i. I've heard rumours [that Sid ate a beetle sandwich].
   ii. What have you heard rumours [that Sid ate _]?
   iii. **Who have you heard rumours [that _ ate a beetle sandwich]?

Presumably the added ungrammaticality of (71iii) over (71ii) arises from fact that the former violates the 'that-trace filter' as well as the complex NP island. Returning now to the examples in (66) - (68), in the worse examples in (66ii), (67ii) and (68ii) the offending subject gap occurs immediately adjacent to the word which introduces the clause: "where", "if" and "and". These words seem to fulfill a similar function as "that" in (70) and (71) in that they serve to introduce a subordinate clause. Informally, then, we might describe all these words as complementisers. In this way, rather than stating a 'that-trace filter' which is specific to a single lexical item, we could instead formulate a new, more generalised constraint which applies to all complementisers (in the broad sense):

72) - The 'complementiser-gap filter' (CGF)*

* C - _  (where C is a variable over all complementisers)

The effect of the CGF is to rule out the occurrence of a gap immediately adjacent to any complementiser. To put it another way (72) prevents the extraction of anything which would otherwise occur next to a complementiser. Note that there is nothing in the CGF which specifically singles out subjects as a special case; subjects will be affected by the CGF incidentally by virtue of the fact that they invariably occur next to the complementisers which introduce subordinate clauses. Bearing this new constraint in mind, we can now account for the contrast between the (i) and (ii) examples in (66)-(68) above. In all these cases, just as with

---

*The CGF really an extension of existing proposals. For example Rizzi (1990) suggests unifying the effects of the 'that-trace' filter with a 'wh-trace' constraint, banning extraction of material immediately adjacent to a WH-word (see also Manzini 1992). As far as I am aware, though, the CGF is unique in applying to all 'complementisers', including conjunctions like 'if', 'after' and "and" as well as WH-words.
(71), the apparent asymmetry between subjects and objects could be said to arise from the fact that extraction of the subject violates the CGF as well as the relevant island constraint. In (66ii) for example the extraction of the subject “who” violates both the WH-island and the CGF, explaining its more severe degree of ungrammaticality in relation to (66i) which only violates the WH-island. So too in (71iii) the extraction of the subject violates the CGF as well as the complex NP island. In (70ii), however, subject extraction violates the CGF alone, and no island.

If, as I have suggested, the apparent extraction asymmetries between subjects and objects are really only a reflex of the CGF and have nothing to do with any inherent property of subjects themselves, then we should of course predict that extraction of a subject which is not adjacent to a complementiser should be exactly parallel to the extraction of an object. Note, for example, the unproblematic extraction of both the subject and object from the ‘weak’ island in (69ii/iii) which, of course, is not introduced by a complementiser. Similarly, as I noted in section 1.2.6, the ‘that-trace filter’, now subsumed by the CGF, does not apply to a complementiser and a subject gap which are not adjacent (see also example (47) above):

73) Who did he say *that* they thought _ were responsible?

More importantly, the CGF also correctly predicts that when not adjacent to a complementiser, extraction of a subject from an island should be no more ungrammatical than the equivalent extraction of an object:

74) i. I've heard rumours [that Sally said Sid ate a beetle sandwich]

ii. */What have you heard rumours [that Sally said Sid ate _ ]?

iii. */Who have you heard rumours [that Sally said _ ate a beetle sandwich]?

75) i. I know [why Rudolf thinks Santa stole the cash].

ii. *What do you know [why Rudolf thinks Santa stole _ ]?

iii. *Who do you know [why Rudolf thinks _ stole the cash]?

According to my intuitions at least, the subject extraction in (74iii) and (75iii) is no worse than the corresponding object extraction in the (74ii) and (75ii) examples. Compare, however, (75iii) with the more ungrammatical example in (76) below where the subject gap occurs immediately after the complementiser:

76) **Who do you know [why _ thinks Santa stole the cash]?
Unlike the sentences in (74iii) and (75iii), the subject extraction here violates the CGF as well as its relevant island constraint. This contrast would of course be very difficult to account for in a theory which attributed the extraction asymmetries shown in (66)-(68), (70) and (71) to inherent properties of subjects and objects.

In this way I believe that an apparent asymmetry between subjects and objects can be reduced to a single incidental constraint, the CGF, which only happens to affect subjects by virtue of their adjacency to complementisers. According to this approach, then, there is nothing inherent about subjects which make them less extractable than objects. However, in spite of its success in accounting for the relevant data, what we still lack is any form of principled explanation for the CGF; why should it be that extraction of an element which is adjacent to a complementiser is worse than one that is not? This is a question to which I will return in section 5.6.2 of the fifth chapter, where I will attempt to put forward a processing-related explanation for the CGF.

1.3.4 - Parasitic gaps revisited

As I noted in section 1.2.7, Parasitic Gaps (PGs) have received a great deal of attention in a variety of theoretical frameworks (Chomsky 1982, 1986-a, Engdahl 1983, Hudson 1984-b, Manzini 1994-a, Pollard and Sag 1994). However, certain aspects of these constructions have, as far as I know, not been fully discussed, and still lack any form of principled explanation within any theory to date. For one thing, we have no explanation of why it is that an extracted argument licences a PG while an extracted adjunct does not:

77) i. Which drugs did Santa take _ without testing _ first?

ii. **Why did Santa take the drugs _ without testing them first _ ?

While (77ii) is fine as a sentence, it evidently cannot be interpreted according to the suggested reading where “why” is supposed to modify both Santa’s taking of the drugs and his not testing them first. Evidently, then, adjunct gaps, unlike argument gaps, do not license PG’s. Any theory which implements a similar displacement mechanism for arguments and adjuncts will have to explain this contrast.

---

10This reasoning is, of course, reminiscent of the discussion of the subject island in section 1.3.2 where I argued that it was similarly the position of subjects rather than any inherent property of them which accounted for their status as islands.
More importantly, perhaps, PG's seem to have an as yet unexplained effect on the "islandhood" of the domains in which they occur. For example the bracketed clause in (78) below is clearly a (tensed) adjunct, and is thus correctly predicted to be an island for extraction:

78) *Which drugs did Santa fall asleep [before he tested _]? 

However as soon a the gap in the bracketed island is made parasitic on another gap in the main clause, then the construction becomes grammatical:

79) Which drugs did Santa take _ [ before he tested _]? 

The apparent immunity of the dependency between "which" and the second gap to the effects of the adjunct island will also have to be accounted for.

Sections 1.3.1 - 1.3.4 above do not by any means present a definitive list of mysterious or poorly-understood factors concerning extraction. It would appear, though, that even relatively little investigation may yield a range of relevant data which seem to have remained unexplored within most syntactic frameworks.

1.4 - Extraction and Syntactic Theory

The complex extraction data reviewed in sections 1.2 and 1.3 pose a unique challenge for any theory of syntax. In particular, a syntactic framework will have to incorporate some means of capturing the long-distance relationships which can arise from displacement, while also balancing the apparently contradictory requirements of extraction in being both seemingly unbounded in certain circumstances and highly constrained in others. Moreover, the complex and intricate range of grammaticality judgements associated with island violations, as well as the various extraction asymmetries between subjects, objects and adjuncts, point to a more subtle and scalar definition of well-formedness than any crude and absolute binary distinction between grammaticality and ungrammaticality.

It is for these reasons that extraction data represent something of a benchmark test against which the adequacy of different syntactic theories has traditionally been evaluated. Certainly any framework which is able to offer a reasonably plausible and elegant account of the complex data reviewed so far deserves to be taken seriously. However, the most important challenge facing a model of syntax is not just to account for the full range of data, but also do
so in a plausible and simple way which might realistically serve as a model of human linguistic knowledge. In other words a theory will have to be both holistic, in accounting for as much data as possible, and realistic in accounting for these data by means of a relatively constrained and learnable (or innate) syntax. These twin requirements of holism and realism correspond roughly to Chomsky’s (1965) criteria of observational and explanatory adequacy, and in some sense balancing these two requirements is the hardest task of any theory of syntax; inevitably the more data one seeks to account for, the harder it will be to resist appealing to ever more specific and stipulative rules and devices, against the spirit of realism.

As I noted in section 1.1, my aim in this thesis is to demonstrate that a dependency-based theory of syntax may offer an adequate account of extraction data, an account which is both holistic and realistic in the senses described above. Having now reviewed the relevant data it may be useful at this point to outline some of the key differences between dependency and phrase structure syntax. I will, in fact, postpone a fuller discussion of these questions until Chapter 3. For now, however, it is worth setting the matter in some sort of context and outlining the central issues involved.

There are two logically possible ways of analysing the syntactic structure of a sentence; as Mel’čuk (1986 p. 13) concludes: “there is no essentially distinct third possibility”. These two methods of analysis can be described as constituency-based and dependency-based. Consider the sentence in (80):

80) My friends have no taste.

According to a constituency-based view of syntax, in analysing (80) we start with the largest possible unit, the sentence, and break it down into successively smaller constituents, or phrases, until we reach the level of words. Thus the sentential constituent marked a in (81i) can be split up into two smaller bracketed phrases in (81ii):

81) i. [My friends have no taste]
   ii. [αβ my friends] [γ have no taste]]

Both the β and γ phrases in (81ii) can be further divided into smaller parts:

82) [αβ [η, my] [γ, friends]] [γ, [ε, have] [η, no taste]]

Matthews (1981) and Siewierska (1988) offer a relatively impartial review of both dependency and phrase structure syntax.

34
Finally the constituent \[ [\text{no taste}] \] can also be split up into its component parts:

83) \[ [\epsilon [\delta [\text{my}]] [\beta [\text{friends}]]] [\gamma [\eta [\text{have}]] [\epsilon [\alpha [\text{no}]] [\delta [\text{taste}]]]] \]

This bracketed system of phrase structure can also be represented by means of a branching tree structure as in (84):

84)

```
    α
   / \   / \
  β   γ  δ  ε
 /     /   /   \
my friends have no taste
```

Eventually, then, we reach the words of a sentence by splitting up the sentence itself into successively smaller units or phrases.

This, in essence, is a phrase structure or constituency-based approach to syntax. Since the early work of Chomsky (1957, 1965, 1970), and others working within his overall theoretical framework, this type of syntactic analysis has come to be identified with the linguistic mainstream, from its inception in the Standard Theory through Principles and Parameters Theory (Chomsky 1981, 1986-b, Chomsky and Lasnik 1993) to the more recent Minimalist Programme (Chomsky 1992, 1995). Other, 'non-mainstream' theories to make use of a phrase structure formalism include GPSG (Gazdar et al. 1985), HPSG (Pollard and Sag 1987, 1994) and LFG (Bresnan 1982), as well as less well-known theories such as Residential Grammar (Binkert 1984). Indeed, so well-established has phrase structure become that one may be forgiven for thinking that it is the only method of syntactic analysis.

However, the other logically possible way to analyse the sentence in (80) is to start with the words themselves, and to seek to link them together directly by simple relations. Thus, for example, a direct relation could be established between “friends” and “my” which could itself be linked to “have”. This type of syntactic analysis is illustrated in (85):

85) \[ \text{my friends} \overset{\text{have}}{\longrightarrow} \text{no taste} \]

This, very basically, is the approach to sentence structure enshrined in dependency theory. According to this view syntactic structure consists entirely of binary relations, or dependencies, between pairs of words. ‘Constituents’ such as sentences and phrases thus have no primitive
significance and are derived instead on the basis of how individual words relate to one another. The structure in (85), for example, is only defined as a sentence because it constitutes a well-formed network of inter-word dependencies. In this way a dependency analysis could be said to take the exact opposite view of syntactic structure from a phrase structure analysis, which, as I described above, works its way down to word level through larger constituent patterns.

The two analyses in (83)/(84) and (85) serve to define two distinct approaches to syntax. As I noted in section 1.1, though, there has been a notable disparity between these two theoretical frameworks with respect to their treatments of extraction phenomena; virtually all constituency-based theories of syntax have relatively well-developed accounts of extraction. Indeed, it is no exaggeration to say that extraction data have played an important role in the shaping and development of the syntactic machinery of many of these theories. For example the transformational component and movement analyses of the Standard Theory (Chomsky 1973, 1977) and later in PP Theory (Chomsky 1981, 1986a, b, Lasnik and Saito 1984, 1992) were to a large degree shaped by the need to accommodate the long distance syntactic relations arising from extraction. Similarly, this requirement has also had an important bearing on the formulation of PS rules and principles of feature transmission in GPSG (Gazdar 1981, Gazdar et al. 1985).

The same, however, can not be said of dependency theories which have generally not paid the same attention to extraction data. As I noted in section 1.1, dependency syntax in its purest form is represented by three contemporary theories, Word Grammar (Hudson 1984a, 1990), Lexicase (Starosta 1988) and Mel’čuk’s (1988) theory. Of these, however, only Word Grammar offers a serious account for extraction data. Mel’čuk (1988) says nothing about the issue at all, while Starosta (1988 p. 257) acknowledges of Lexicase:

“Other areas of lexicase are in a flux or are only beginning to be explored ... these include in particular the structure of indirect and direct wh-questions, relative clauses ... topicalised sentences...”

These, of course, are precisely the constructions relevant to extraction, as discussed in sections 1.2.1 and 1.2.2. It is this imbalance between constituency and dependency-based theories with respect to extraction data which, in a small way, I hope to redress in the rest of this thesis.

---

12However, it is fair to say that within the Minimalist Programme (Chomsky 1993, 1995), the latest phase of PP Theory’s development, issues relating to extraction have not yet received a great deal of attention.
1.5 - THESIS OUTLINE

The remainder of this thesis is organised as follows: Chapter 2 will offer an overview of the Principles and Parameters (PP) Theory account of extraction. PP Theory constitutes part of the constituency-based linguistic mainstream, which describes extraction phenomena in terms of syntactic movement between deep and surface levels of structure. It is fair to say that PP Theory offers the most comprehensive treatment of extraction data of any syntactic framework, and thus I believe that the discussion in the second chapter will provide an essential context within which the dependency-based approaches to these data can be properly evaluated in subsequent chapters.

In chapter 3 I will outline the theory of Word Grammar (WG). WG is the best-developed modern dependency-based theory of syntax, and the only one to have formulated a serious account for extraction phenomena. WG is a monostratal framework which does not recognise any equivalent of syntactic movement, and I will pay particular attention to how these broader theoretical assumptions define a specific approach to extraction data. Although WG is remarkably successful in formulating a relatively straightforward account, I will show that it is unable to match PP Theory in the range of data that it covers.

Chapter 4 presents my own interpretation of a monostratal dependency grammar, an outline proposal which I refer to as Licensing Grammar (LG). LG seeks to redefine dependency relations as licensing relations. I will argue that this points to a more constrained, 'transparent' syntax which, unlike most other dependency theories, does not necessarily have to recognise distinct grammatical relations.

In the fifth chapter I will examine how the theory of LG may serve as the basis of an account for the extraction data discussed in the first three chapters. I suggest that a purely structural approach to these data is inappropriate, and that only by incorporating processing-related considerations might we have a realistic chance of capturing the variable nature of island effects. I will outline a parsing system for LG and show how the operation of this parser may serve to constrain certain syntactic principles. I will argue that this collaboration between dependency syntax and processing allows for the articulation of a flexible account of extraction which can be shown to enjoy some important advantages over both the PP Theory and the WG treatment of the same data. In this way I hope to establish that a dependency-based theory of syntax is indeed capable of serving as the basis of a plausible and effective account of extraction.
CHAPTER 2

PRINCIPLES AND PARAMETERS THEORY AND EXTRACTION

2.1 - Introduction

Although the main topic of this thesis is dependency grammar and its capacity to account for extraction data, in this chapter I will first offer a review of Principles and Parameters Theory, and examine its own account of extraction. As I pointed out in the introduction to the first chapter, Principles and Parameters Theory (henceforth PP Theory) represents the focal point of the constituency-based linguistic mainstream. As such, it has made more progress than other theories in offering a principled and adequate account of extraction phenomena, and in many ways the PP account of extraction sets the standards against which other theories must be judged. As Hudson (1993 p. 337) acknowledges:

"There is surely no doubt that far more progress has been made over the past few decades within the framework of constituent structure than among the various strands of dependency theory - far more sophisticated theories of constituent structure have been developed and tested against far more data. Advocates of dependency theory must address the same range of problems and accept the same standards of success as these other theories."

In this chapter, then, I will seek to outline some of the crucial aspects of PP Theory's treatment of extraction and evaluate its 'standards of success' in accounting for the complex body of data discussed in the previous chapter. This review will then provide a context for my later exploration of how a dependency-based framework might account for the same data. What follows is not intended as a general introduction to PP Theory, nor an exhaustive survey of the theory's various accounts of extraction data; there already exist many such introductions and summaries, notably Chomsky (1986-b), Horrocks (1987), Radford (1988), Haegeman (1991), and Chomsky and Lasnik (1993). Instead this section is intended as a guided tour and critical review of some key aspects of the broad PP approach to the data discussed in the previous chapter, and an exploration of how this account relates to the theory as a whole. In section 2.2 I will outline some fundamental characteristics of PP Theory, notably its use of phrase structure and movement analyses. Then in sections 2.3 and 2.4 I will go on to describe how these components underpin a specific approach to extraction data.

---

1I will use Principles and Parameters (PP) as something of a generic term covering various stages in the theory's development including the Standard Theory and the Extended Standard Theory.
2.2 - Some Fundamental Issues in Principles and Parameters Theory

It is in fact somewhat misleading to refer to any single PP account of extraction. PP Theory is a general theoretical framework within which numerous people have put forward different explanations for all sorts of linguistic phenomena. Extraction is no exception, and within PP Theory alone we find Chomsky (1981, 1986-a), Kayne (1981, 1984), Lasnik and Saito (1984, 1992), Aoun (1985), Rizzi (1982, 1990), Cinque (1991) and Manzini (1992, 1994-b, c), amongst others, all dealing to a greater or lesser extent with extraction and offering different ideas as to how to account for constraints on its operation. Naturally, though, all treatments of these phenomena advanced within the broad PP framework share a substantial core of common assumptions and technical devices. More particularly, these accounts all utilise the two fundamental components which in many ways define PP as a theory: a branching phase structure configuration to represent the syntactic structure of sentences, and the notion of movement or derivation, by which the surface syntactic form of a sentence is derived from an underlying or 'deep' structure. In order to afford a fuller understanding of the PP Theory approach to extraction I will briefly examine each of these issues in turn.

2.2.1 - Constituency and clause structure

Figure (1) below shows the type of branching configuration used as a formal representation of sentence structure in PP Theory. CP and IP are functional projections associated with the lexically-headed VP constituent. The I head of the IP projection bears an abstract inflectional element commonly realised as morphological inflection on the verb, and the C head of the CP projection houses the complementiser in embedded constructions:

As I pointed out in section 1.4 of the first chapter, (1) exemplifies a phrase structure-based approach to syntax according to which larger phrases are broken down into successively smaller units until reaching word level. According to governing principles of phrase structure
(or X-bar theory as it is known in PP), each configurational structure is comprised of identical interlocking modules or phrases of the type shown in (2) (Jackendoff 1977, Stowell 1981):

2) 

Each of these modules constitutes a maximal projection (XP) of a head (X). An XP consists of two elements, X' (an intermediate projection of X) and the specifier of X (spec). The intermediate X' projection is itself a constituent comprising X (the head of the projection) and a complement (comp). A general constraint of Modifier Maximality (Jackendoff 1977) determines that the specifier and the complement of a head will themselves be maximal projections. In this way each XP constituent will be housed within another constituent, YP, either as the specifier or complement of Y, the only exception to this being the matrix CP, which is clearly not contained within any larger constituent. According to this conception of syntax, words, which correspond to the terminal (X^0) nodes of the tree, are linked together by their participation in this system of phrase structure.

Within this sort of clausal structure, the occurrence of a nominal constituent - NP - is licensed by Case assignment (Chomsky and Lasnik 1977, Chomsky 1986-b); the lexical V^0 head assigns Accusative Case to its complement, the object NP, while the functional I^0 head assigns Nominative Case to its specifier, the subject NP. In this way subjects and objects can be defined according to their position within the phrase structure, and in PP Theory subject and object relations have no primitive significance of their own (Marantz 1984, Burzio 1986, Speas 1990). Case-assignment by V^0 or I^0 is subject to a structural relationship of government, about which I will say more later. Thus a structure for the sentence in (3i) might be (3ii), with certain details omitted for the time being:

3) i. Noam plays the banjo.

ii. 

Thus a structure for the sentence in (3i) might be (3ii), with certain details omitted for the time being:
Following Chomsky (1970) and Stowell (1981), the internal structure of NPs is identical to the schema in (2), each NP constituting the maximal projection of its N₀ head.

One further possibility allowed within this system of phrase structure is adjunction; an element can be adjoined to a constituent forming a copy of that constituent. Consider the structure in (4ii):

4) i. Noam plays the banjo badly.

Here the adjunct "yesterday" is adjoined to IP, creating a copy of the original IP. The fact that adjuncts are adjoined in this way, and are thus structurally further removed from their head, captures the intuition that they are somehow less involved with the head than complements and specifiers. Assuming that adjuncts will always be adjoined to a maximal projection (Chomsky 1986-a, Kayne 1994), like subjects and objects they will be defined on the basis of their structurally-defined position.

This configurational sentence structure has two major implications that are relevant to extraction. Firstly, the schema in (1), (3ii) and (4ii) allows for the expression of certain geometrical relations holding between different elements in the tree. Thus, for example, it is possible to postulate entirely structurally-defined relationships such as c-command (Langacker 1969, Reinhart 1976, Chomsky 1981), m-command (Chomsky 1986-a) and more restrictive variants thereof, such as Government (Chomsky 1981, Aoun and Sportiche 1981), L-marking (Chomsky 1986-a) and g-marking (Manzini 1992). I will go on to describe some of these relations in greater detail later on. For the moment, however, the important point is that all these syntactic relationships, many of which play a crucial role in PP Theory-based accounts

Following Abney’s (1987) proposal, nominal constituents are usually described as determiner phrases (DPs) headed by a determiner which takes a conventional NP as its complement. See also Hudson (1984-a) and Horrocks and Stavrou (1987).
of extraction, are inseparably bound up with the phrase structure illustrated in (1), (3ii) and (4ii), and would remain inexpressible without such a configuration.

The other significant implication that constituent structure has for extraction is that it allows for, indeed it demands, the recognition of distinct structurally-defined positions. As I pointed out above, the recursive nature of X-bar theory allows us to recognise a set of positions such as complement of \( V^0 \), and specifier of \( I^0 \). Indeed we may abstract away from categorial information entirely, and express generalisations pertaining generally to specifiers, complements and adjoined positions. In this way we can refer to positions within a structure without necessarily taking account of their content; any head will have a specifier and complement ‘slot’ irrespective of whether or not these positions actually contain any overt material. By their very nature the XP and X’ constituents only exist insomuch as they make complement and specifier positions available. This last point is crucial for PP Theory’s approach to extraction, since these structurally-defined positions allow for a coherent and reasonably principled theory of empty categories, invisible constituents whose presence can be (partly) justified on the basis of there being a position available for them to occupy. These empty categories play a key role in all PP-related accounts of extraction phenomena, and will be discussed in greater detail below.

2.2.2 - Movement

Consider the sentence in (5) below:

5) What does Noam play?

Here “what” functions as the object of “play”. Objects in English invariably follow the verb as in the sentence “Noam plays the banjo”. One way of approaching the sentence in (5) is thus to say that “what” actually originates in a typical object position, after the verb, and is then transplanted to the beginning of the sentence. This, in a nutshell, is PP Theory’s approach to extraction phenomena (see, for example, Chomsky and Lasnik 1993); a surface syntactic structure such as that in (5) will have the “deep” representation shown in (6), where the WH-word “what” occurs in its canonical post-verbal object position:

6) Noam plays what?

Within earlier versions of the PP model, such as the Standard Theory, this ‘deep structure’ (D-Structure) would then be mapped onto the surface form (S-Structure) shown in (5) by a
specific transformation, a rule specifying the movement of the WH-word from its post-verbal position to its surface clause-initial position. See Chomsky (1973) for an example of how such a rule might be formulated.

From its very inception, the Standard Theory, the forerunner of PP Theory, recognised movement of this type as an important part of syntax, and consequently incorporated a system of transformational rules, each of which specified a similar individual mapping between D-structure and S-structure. This system of rules was known as the transformational component of the grammar; together with a system of phrase structure rules (PS rules), which was responsible for generating the configurational patterns discussed above, the transformational component constituted the syntactic basis of the theory, and, as such, part of a speaker’s linguistic competence (Chomsky 1965).

In spite of the central role subsequently played by transformations in mainstream syntactic theory, it is important to remember that originally the transformational component of the Standard Theory was introduced not on the basis of its intrinsic desirability, but rather as a means of compensating for the perceived inadequacy of the generative capacity of PS rules (Chomsky 1957 ch. 5). Quite simply PS rules of the type discussed in Chomsky (1957, 1965) were unable to generate anything other than relatively simple structures, with all elements in their canonical position such as the D-Structure in (6). The transformational component then operated on the output of these PS rules yielding more complex S-Structures involving, for example, long-distance dependencies between elements. One way of viewing transformations is thus to see them as reducing complex long distance relationships, commonly encountered in surface syntactic forms, to simpler local ones which can be more adequately handled by PS Rules. To take a specific example, consider the following sentences:

7) i. Which syntactician do you think (x) is mad?
   ii. Which syntacticians do you think (x) are mad?

---

3 Originally only some of the rules in the transformational component actually involved movement; others allowed for the deletion, copying or alteration of material (Chomsky 1973).

4 It is worth noting at this point, however, that the inadequacy of phrase structure rules was only perceived. PS rules can in fact be elaborated so as to account for more complex data, including extraction (Gazdar 1981, Gazdar et al. 1985, Pollard and Sag 1987, 1994).
Evidently here the form of the WH-phrase is responsible for controlling agreement on the embedded verb “is”/“are”. Limitations of early PS rules dictated that agreement associations such as this, in common with most other syntactic relationships, had to be treated in a local manner; for example, an NP might only control agreement on a neighbouring verb. Thus by suggesting that in (7) above the WH-phrase originates in the position marked (x), at D-Structure a suitable local relation could be maintained between it and the agreeing form of the verb, irrespective of how far the two constituents find themselves from each other at S-Structure. Horrocks (1987 p. 48) summarises the argument:

“Every argument that was ever advanced in the standard-theory literature for introducing deep structures and some transformational rule such as wh-movement ultimately reduces to the claim that the formalism solves the problem of enforcing a complex or ‘long distance’ dependency by reducing it to a simple or ‘local’ one, thereby allowing the rules stating the nature of the dependency to operate on adjacent items in every case”.

Another argument advanced in favour of D-Structure, and hence, by extension, in favour of transformations such as movement, was that originally D-Structures were considered to be the most suitable constructs from which semantic information could be read off, providing as they do an almost direct syntactic representation of predicate argument structure (Jackendoff 1972).

2.3 - A Movement-Based Approach to Extraction

2.3.1 - Move-α

In many ways the central role played by transformations has been a distinguishing characteristic of PP Theory and its forbears, and the one most frequently challenged by rival theories. See, for example, Hudson (1976, 1984-a), Perlmutter (1983), Binkert (1984), Gazdar et al. (1985), and Pollard and Sag (1987, 1994). Nevertheless, within PP Theory extraction has been consistently described in terms of syntactic movement, part of a broader derivational process mapping D-Structures onto surface syntactic forms. Given that this derivational component has been such a cornerstone of PP Theory from its very origins, it is not surprising that the transformational mechanism itself should have undergone radical revaluation and reformulation within the course of the theory’s development.

In fact individual transformational rules specifying, for example, the movement of a WH-word within a given context have been abandoned within PP, having been subsumed by a more general derivational operation of Move-α (Chomsky and Lasnik 1977, Stowell 1981, Chomsky 1981). The operation of Move-α effectively allows any element within a D-Structure
to move anywhere. It is then up to separate, independent components of the theory to constrain this general ‘wildcard’ operation. Thus principles of Case Theory, X-bar Theory and Binding Theory, for example, might all conspire to restrict significantly the potential freedom of a word to move anywhere, allowing it perhaps the option of moving to just one position within a structure (Chomsky 1981, 1986-b, Chomsky and Lasnik 1993). Certainly the role of transformations and the history of their development within PP Theory is far too broad an issue to be adequately examined here; clear accounts can be found in Lasnik and Uriagereka (1985) and Horrocks (1987). The important point here, though, is that in spite of these developments, PP Theory still describes extraction in terms of syntactic movement; the operation of move-α isolates an element at D-Structure and moves it to another position which it occupies at S-Structure.

With respect to the configurational phrase structure described in section 2.2.1 above, WH-extraction involves the movement of a WH-word from its original position at D-Structure to the specifier position of CP, commonly referred to as [spec, CP] - see examples (9) and (10) below. [spec, CP] is a non-argument position; it consistently remains empty at D-Structure, its main function being to house displaced material, such as WH-words and topicalised elements, at S-Structure. For this reason, then, WH-movement is generally described as A̅-movement, movement to a non-argument (A̅) position (Cinque 1991).

2.3.2 - Empty categories and chains

Originally, according to the early Standard Theory view, movement took place in a relatively simple fashion, an element vacating its D-Structure position and moving elsewhere. For a variety of reasons, which are not of immediate relevance here, it became increasingly apparent, however, that this was an inadequate conceptualisation of movement. For one thing the moved element quite literally disappeared ‘without trace’ from its original position; this often had the effect of radically altering the syntactic structure of the source environment, violating well-motivated principles of structure preservation (Emonds 1976) and selectional uniformity - the Projection Principle as described by Chomsky (1981). In response to problems of this sort it was suggested that a moved element might leave behind a non-overt copy of itself, or trace, in the position it vacated at D-Structure, analogous, perhaps, to a footprint.

Lasnik and Saito (1984) develop a yet more general operation of Affect-α, subsuming not only cases of movement, but also insertion and deletion.
(Fiengo 1974, Chomsky 1975, Lightfoot 1977). In this way the trace, a full NP (though one lacking phonological content), would remain in the source position, not only allowing structure to be preserved, but also keeping track of an element's original D-Structure position at S-Structure. Thus a simple surface representation of the sentence in (5) above might look something like (8), where 't' is a trace, a non-overt copy of "what". According to representational conventions, the referential co-identity of "whom" and its trace is indicated by means of co-indexation:

8) What does Noam play?

However, certain properties of "what" in (8) can actually be better expressed in terms of its trace. For example it is through the position of t that "what" is defined as an argument of "play"; after all, the trace occupies the complement position of the verb where it will be assigned Case and a θ-role (Chomsky 1981). For this reason it is convenient to think of "what" and its trace as being linked by a chain <what, ... t>. Thus any Case or θ-role assigned to t can then be transmitted to "what" through this syntactic chain. As I will discuss in greater detail below, the syntactic chains linking a displaced element and its trace(s) play an important part in PP Theory and its treatment of extraction data.

Returning now to the configurational sentence structure first illustrated in (1) of section 2.2.1, the WH-word "what" in (8) above vacates the complement position it occupies at D-Structure and moves up to [spec,CP] leaving a co-indexed trace of itself behind. The sentence's D-Structure and corresponding S-Structure would thus be represented respectively as (9) and (10) below:

9)  ```
    CP
   /   \
spec 
   /   \  
CP    C'
  /    \   
P   IP
   / \  
NP  I'  
   / \   
  V  VP 
 /  /  
Noam plays the banjo
```
Note that the C⁰ position, which is empty at D-Structure, serves to house the auxiliary verb “did” in (10) which would otherwise occur in the I⁰ position. However the absence of ‘auxiliary inversion’ in embedded interrogatives and topic constructions means that in these cases the C⁰ will remain empty.

The introduction of non-overt traces has some far-reaching consequences for the concept of syntactic movement. Most importantly, perhaps, with the existence of traces at S-Structure, the entire history of a moved element’s progress during a syntactic derivation is encoded at a single level of representation by virtue of the chain connecting it with its trace(s). This consequently opens the way to expressing movement constraints as representational conventions, such as well-formedness conditions on chains, holding at S-Structure (Koster 1978, 1986). The logical conclusion of this argument is that ultimately D-Structure becomes redundant, serving only to reduplicate information visible at S-Structure. Assuming that all constraints and conventions of D-Structure can be adequately expressed elsewhere in the grammar, the strong implication of this is that D-Structure as a separate level of representation could be eliminated altogether. This is in fact the position adopted within the Minimalist Programme, the latest incarnation of PP Theory (Chomsky 1992, 1995, Brody 1995).

For now, however, the important point is that PP Theory’s conception of movement allows certain important generalisations concerning extraction phenomena to be expressed in terms of chains and their well-formedness. Notice in particular that the concept of chain is intimately bound up with the configurational sentence structure discussed in section 2.2.1; as pointed out there, the notion of an empty category - which is central to chain-formation - only really makes sense within the context of a branching configurational structure which makes available distinct structural positions for these empty categories to occupy. In addition to this, well-formedness conditions on chains, and hence constraints on movement operations, can be
conveniently expressed in terms of configurationally-defined structural relationships obtaining between the two (or more) elements of a chain. One such structural relationship is c-command (Reinhart 1976, Chomsky 1981), stated informally in (11):

11) C-Command

An element X will c-command its sister Y and all elements dominated by Y

I assume here familiarity with the notions of “sister” and “domination”. With respect to chains, then, we can impose a condition that the antecedent - the moved element - must c-command its trace - the root of the chain (Chomsky 1986-b). The chain <what,... tj> in (10), for example, meets this criterion; the antecedent “what” c-commands its sister C’ and everything that C’ dominates, including the trace tj. Evidently this relationship is asymmetrical in that the trace cannot c-command its antecedent.

The c-command condition on chains, guarantees that the antecedent will always occur ‘higher up’ in the tree than its trace. This is important since it also makes a prediction relating to the ordering of elements in a chain. In a simple right-branching phrase structure of the type illustrated so far, if an element X c-commands another element Y then this entails that X will also precede Y (Kayne 1994). This will therefore explain the fact that in English extraction is always from right to left, since the ‘extractee’ will have to c-command, and thus precede, its trace. Thus rather than having to state the leftward nature of extraction as a primitive fact of English, we can instead derive it as a theorem from the configurational structure of the language coupled with the general c-command condition on chains. This offers one insight into the potentially intimate relationship between phrase structure and movement, the two key theoretical components of PP Theory discussed in sections 2.2.1 and 2.2.2.

2.3.3 - Logical Form and ‘non-overt’ movement

As we have seen, the operation of Move-α is a syntactic derivational procedure which maps D-Structures onto S-Structures. Naturally, since S-Structure is taken to be the surface syntactic form of a sentence, it is at this level of representation that the structure will be subject to phonetic interpretation. To put it simply, a sentence will be heard in the form in which it

6Of course in a left-branching (or head-final) language the c-command condition would predict that the extractee followed its trace. As I noted in section 1.2.1 of the previous chapter, however, systematic rightward extraction is unattested among the world’s languages, and thus the absence of extraction in head-final languages will still have to be stipulated (Chomsky 1973, 1977).
exists at S-Structure. In the surface forms in (8) and (10) above, for example, we ‘hear’ the
WH-word “what” in a sentence-initial position, after Move-α has placed it there.

Once one allows for distinct levels of linguistic representation over and above what is
actually heard in speech - as with D-Structure - and one has in place a derivational mechanism
such as Move-α, allowing for a principled mapping between these levels, there is no reason
why this same derivational mechanism should not continue beyond S-Structure, yielding still
further levels of representation which, again, need not be subject to phonetic interpretation.
This possibility is exploited in PP Theory; the operation of Move-α is seen to continue beyond
surface syntax, deriving one further level of representation - Logical Form (LF). The principle
motivation behind this additional level is that LF, abstracted away from phonetic “reality”, may
serve as a more accurate representation of the logico-scopal properties of a sentence than the
and Sportiche 1981). In (12i) below, for example, the quantifier “everyone” logically takes
scope over the entire structure, and the sentence’s representation in Predicate Calculus would
be something like (12ii) (Cann 1993 ch. 6):

12) i. Minimalism confuses everyone
    ii. ∀x C (m, x) where x = person, m = Minimalism and C = confuse

A simplified LF of the sentence in (12i) might thus approximate (12ii), as illustrated in (13):

13) Everyone, Minimalism confuses t,

Here, then, after the sentence reaches S-Structure, yielding the surface form encountered in
(12i), Move-α applies to the quantifier “everyone” and moves it to some position to the left,
higher up the tree, where at LF it may take appropriate scope over the whole sentence.
Naturally, though, this further derivation is not ‘heard’.

It has been suggested that WH-words are quantifiers which themselves must take scope
over their sentence (Chomsky 1975, 1977 Higginbotham and May 1981, May 1985). If this
is correct, then clearly by LF all WH-words will have to be in a position from which they can
take scope over their clause, even if in S-Structure they occur in situ, as in (14):

14) Sid ate what?

As I pointed out in Chapter 1, the construction in (14) is marked in English, and could only
be uttered within a specific discourse context, expressing surprise or incredulity. Recall,
though, that in languages which do not allow extraction, such as Japanese and Turkish, *in situ* constructions such as (14) are the norm. If then by LF all WH-words must satisfy their requisite scope-taking properties, then those which occur *in situ* at S-Structure will have to move covertly to a sentence-initial position at LF. In this way syntax of all WH-interrogatives, whether *in situ* or *ex situ* at S-Structure, can be captured in a uniform manner. The LF representation for (14) LF will thus be (15):

15) \( \text{What, Sid ate } t_i. \)

In this way according to PP Theory all WH-words in all languages will be moved; being quantifiers, by LF they will have to occupy a clause-initial position such as \([\text{spec,CP}]\) so as to satisfy their relevant scopal properties. The only question concerns whether this movement takes place overtly, between D-Structure and S-Structure, as it usually does in English, or covertly, between S-Structure and LF, as in Japanese and Turkish. A schematic representation of this system of derivation is shown in (16)^7:

16) D-Structure
    ↓
    Move-\(\alpha\) (overt)
    ↓
    S-Structure \(\rightarrow\) PF (phonetic interpretation)
    ↓
    Move-\(\alpha\) (covert)
    ↓
    Logical Form

Evidence for the covert application of Move-\(\alpha\) after S-Structure is sometimes thought to come from the so called ‘superiority effects’ illustrated in (17) below:

17) i. Wally wore a fez to go to the party
    ii. Why\(j\) did Wally wear what \(t_j\) ?
    iii. *What\(j\) did Wally wear \(t_i\), why?

According to Chomsky (1981) and Aoun, Hornstein and Sportiche (1981) the ungrammaticality of (17iii) is said to arise from the illicit nature of the covert movement of “why” after S-Structure to its LF position, represented in (18):

18) *Why\(j\) [what\(i\), did Wally wear \(t_i, t_j\)]

^7Brody (1992) examines some theoretical implications of this system of grammatical derivation.
The movement of "why" here is said to be blocked by the presence of "what" in [spec,CP] position (see section 2.4.1 below). In contrast, the covert movement of "what" in (17ii) is supposed to be unproblematic, thus accounting for the grammaticality of this sentence.

Clearly arguments such as these open up a whole new dimension in the research of extraction data since the behaviour of in situ elements such as those in (17) now has to be accommodated within the broader account of overt displacement. It is important to remember, however, that the notion of 'covert movement' is entirely specific to PP Theory with its derivational mechanism of Move-α and distinct representational levels of S-Structure and LF, and thus the contrast between (17ii) and (iii) will not qualify as extraction data in other theories which 'only' deal with overt, observable data. For this reason I will not discuss examples such as (17) further in this thesis*

2.4 - Constraints on Extraction

The combined properties of phrase structure, movement and trace theory outlined in the previous section allow for the articulation of a complex and intricate account of extraction phenomena. As I pointed out above, the general PP Theory conception of extraction involves a WH-word moving to [spec,CP], leaving a non-overt copy of itself, a trace, in its source position. The moved element and its trace then constitute a syntactic chain, encoding at one level of representation, be that S-Structure or LF, the derivational history of that element. So far, though, this syntactic movement, a case of the more general derivational operation of Move-α, has been presented as being largely unconstrained, subject only to the c-command condition on chains discussed in section 2.3.2. Evidently in itself this condition is not a sufficiently restrictive constraint on movement; according to the formulation of c-command in (11), an element may move indefinitely higher up a tree and still c-command its trace. The 'island' data in the previous chapter clearly reveal the extent to which extraction is constrained, however, even when the distance between an extractee and its trace is comparatively small. For this reason it will be necessary to supplement the c-command condition on chains with

---

*There is of course still a question to be answered as to why (17iii), unlike (17ii), is ungrammatical. However, this contrast will have to be attributed to some general property of adjuncts rather than any putative constraint on their 'covert mobility'. As I noted in section 1.2.1 of the first chapter, in situ WH-adjuncts tend to be problematic even outside WH-contexts:

1. *Wally wore a fez why?
2. *"Noam plays the banjo how?"
restrictions on the quantity and nature of material intervening between the displaced WH-element and its trace(s).

2.4.1 - Subjacency

Consider some of the islands described in the previous chapter with their respective syntactic structures:

19) *What does [chasing _ ] tire Mr. Neville? (subject island)

20) *What did Sid fall ill [because he had eaten _ ]? (adjunct island)
21)  "What do you believe [the story that Sid ate _ ]? (complex NP island)

22)  "What do you know [why Wally wore _ ]? (WH-island)
In all of the above examples extraction of the WH-word gives rise to ungrammaticality. The first systematic survey of such examples is Ross (1967). Writing within the somewhat restricted scope of the early Standard Theory, his account effectively amounts to an inventory of individual island constraints. According to such an approach, then, each of these islands would be listed as a discrete constraint on the operation of the specific transformation moving a WH-word to the clause-initial position. Thus example (19) would be described as a violation of the Sentential Subject Constraint (or subject island), a specific ban on the extraction of an element from a subject domain. Similarly (20) and (21) violate the adjunct island and complex Noun Phrase island respectively, while (22) constitutes a violation of the WH-island. Each of these island constraints remain distinct and no attempt was made to offer a more integrated, holistic account of these phenomena. Of course, the discrete statement of each island constraint is stipulative and makes no prediction as to what could or could not conceivably constitute such a domain. Why, for example, should subject clauses be islands but not object clauses?

Within later versions of the theory efforts were made to offer a more principled account of these data whereby each island could be subsumed by more general properties constraining the application of extraction, and indeed transformations in general. Chomsky (1973), for example, formulates various general restrictions constraining the operation of transformations, such as the Specified Subject Condition, the Tensed Sentence Condition and Subjacency. Of these, Subjacency is the most relevant to the current discussion, since it serves to derive the effects of at least some of the island phenomena discussed above, namely the complex NP island and the WH island.

Subjacency is a wide-ranging constraint which prohibits the operation of any transformation, including movement, across more than one cyclic node at any one time, a cyclic node, or Bounding Node, being IP (originally S) and NP. Thus any single movement crossing, say, an NP and an IP or two IPs will violate Subjacency and will result in an

---

9 Ross actually describes four domains from which WH-extraction is restricted, the Complex Noun Phrase Constraint, the Sentential Subject Constraint, the Co-ordinate Structure Constraint (ch. 3) and the factive island (ch. 6). He does not discuss constraints on extraction from WH-clauses nor adjunct domains. A further constraint on extraction discussed by Ross, the left-branch condition, has been rendered obsolete by later developments in the theory (Abney 1987).

10 In fact as the theory stood in 1973, the formulation of Subjacency allowed the subject island to be subsumed too, since subject clauses, unlike complement clauses, were thought to be embedded inside an NP. Thus movement out of a subject clause would cross both IP and NP. Later, however, this analysis was abandoned, and the subject island had to be explained by other means (Stowell 1981, Huang 1982). See below.
ungrammatical structure. It is of course relatively easy to formulate Subjacency in purely representational terms, as a well-formedness condition on chains; the two elements of a chain - the antecedent and trace - may not be separated by more than one bounding node. In fact originally CP (S') rather than IP (S) was taken to be a bounding node (Chomsky 1973, 1977, 1981). Following Sportiche (1981) and Rizzi (1980, 1982), however, the choice between IP and CP was said to be parameterised between different languages, with English selecting IP as its option. Thus the ungrammaticality of the complex NP island violation in (21) can be attributed to the distance between the extracted WH-word “what” and its trace which spans three bounding nodes, two IPs and a NP. Similarly in the WH-island in (22) the moved WH-word “what” is separated from its trace by two IPs. In this way both of these examples will violate Subjacency.

Note, though, that this conception of Subjacency requires the postulation of something like an escape hatch for most embedded clause constructions. Consider (23):

23) Whati did [ip Noam tell you [cp that [ip he could play t_i]]?]

Here the perfectly grammatical extraction of “what” is seen to cross two IPs, which should give rise to a comparable degree of ungrammaticality to the examples in (21) and (22) above. In order to overcome problems such as this Chomsky (1973, 1977) suggests an analysis involving intermediate traces. According to this proposal the movement of “what” in (23) to the matrix [spec,CP] position does not take place in one go. Instead the WH-word moves first to the [spec,CP] position of the embedded clause, leaving there a second co-indexed trace (t_i') before moving on to the matrix [spec,CP] position:

24) What, did [ip Noam tell you [cp t_i' that [ip he could play t_i]]?]

Each of the two distinct movements in (24) crosses only one IP, the [spec,CP] position of the embedded clause serving as an escape hatch for the otherwise illicit movement of “what” from the root t_i position. Note also that in (22), a WH-island violation, this option is not available to the itinerant WH-word; the [spec,CP] position of the embedded clause is already filled by another WH-word “why”, thus forcing “what” to move directly to the matrix [spec,CP] in one journey, crossing two IP nodes\(^\text{11}\). This consequently results in a Subjacency violation. So too in (21), a violation of the complex NP island, even the possibility of an

\(^{11}\)A principle known as ‘successive cyclicity’ prevents “what” from passing through the embedded [spec,CP] position before the arrival there of “why” (Chomsky 1973, Pullum 1979).
intermediate trace in the embedded [spec,CP] position fails to rescue the structure since the
distance between any such intermediate trace and the matrix [spec,CP] position still spans two
bounding nodes, NP and the matrix IP, resulting once again in a Subjacency violation.

2.4.2 - Huang's Constraint on Extraction Domains (CED)

Clearly, the ungrammatical subject island violation in (19) and the adjunct island
violation in (20) above cannot be excluded in terms of Subjacency as formulated in Chomsky
(1973). Consider once again the partial structural representations of these two sentences:

25) *What does [chasing _ ] tire Mr. Neville?

26) *What did Sid fall ill [because he had eaten _ ]?

This time I have included the intermediate traces in the embedded [spec,CP], since evidently
in these examples there is no overt material occupying this position so as to prevent such a
trace from occurring there. Clearly, then, in both of the above examples, neither movement
from the source position (t_i) to the intermediate t'_i in the embedded [spec,CP], nor movement
from this position to the matrix [spec,CP] crosses more than one bounding node, and thus the
island effects associated with these sentences have to be explained in terms of something other
than Subjacency.
It is in order to account for examples of this sort that Huang (1982) formulates a
Constraint on Extraction Domains (CED) which effectively subsumes the otherwise discrete
subject and adjunct islands, as well as the relative clause island. The CED, in effect, bars
extraction from a domain that is not properly governed. Proper government is a structurally-
defined relation based on government, a more localised version of c-command, stated in (27)
(Chomsky 1981, Aoun and Sportiche 1981):

(27) Government

A governs B if A c-commands B and there is no barrier between A and B.\(^\text{12}\)

The precise definition of government has been one of the major preoccupations of PP Theory
throughout its development, and I do not intend to include a full discussion of the issue here
1991). Certainly, the definition in (27) is a simplification; for our purposes, though, it suffices
to say that X will govern every YP within its own maximal projection (XP), though not
constituents adjoined to XP, and will also generally govern the head and specifier of any YP
within XP (Manzini 1992). Thus in (26) above, for example, the C\(^0\) head of the matrix clause
will govern everything in its projection (CP), the NP “what” in its specifier position and the
IP in its complement position. C\(^0\) will also govern the head and specifier of this NP and IP,
including the subject “Sid” in [spec,IP]. Similarly the entire NP “what” in the matrix [spec,CP]
position in (26) will govern everything else in CP, the projection in which it is housed,
including the C\(^0\) head and IP. In this way, referring back to the definition in (27), each maximal
projection, XP, will constitute a ‘barrier’ for its complement, preventing that complement from
being governed by anything outside XP. Thus in (26) the IP, for example, serves as a barrier
preventing its complement VP from being governed by the C\(^0\) head.

Proper government, on which Huang (1982) bases his CED, is a more restricted
version of government:

(28) Proper Government.

A properly governs B if and only if A governs B and ...

i) A is a lexical category, or

ii) A is co-indexed with B.

\(^{12}\)Chomsky (1986-a) defines government in terms of m-command rather than c-command, the main
difference being that a head will m-command but not c-command its specifier.
Neither the subject clause in (25) nor the adjunct clause in (26) fulfill these criteria of proper government, and are thus predicted by the CED to bar extraction; these clauses have no governing co-indexed antecedent and neither are they governed by a lexical head. Thus in (25) displacement of "what" from the subject CP falls foul of the CED since this CP is only governed by the non-lexical I° head. The same applies to the adjunct CP in (26), which, being adjoined to IP, will not be properly governed by a lexical head either. In fact the definition in (28) guarantees that apart from elements which have a co-indexed antecedent, the only domains to be properly governed, and thus allow extraction, are the complements (and possibly specifiers) of lexical categories such as N, P and V. Thus the CED effectively prevents extraction out of anything other than a complement clause.

Between them Subjacency and the CED account for the majority of island phenomena discussed in the first chapter. Whereas the CED prohibits extraction out of anything other than a complement clause, thus explaining the subject and adjunct islands, as well as the relative clause island, Subjacency rules out extraction from precisely the two complement domains from which extraction is forbidden, clausal complements of nouns, in the case of the complex NP island, and WH-clausal complements of verbs, in the case of the WH-island. In this way the CED and Subjacency complement one another in their respective areas of application.

2.4.3 - The ECP, extraction asymmetries and 'that-trace' effects

Recall from the first chapter that constraints on extraction coincide with certain asymmetries concerning the extractee itself. More specifically, the displacement of adjuncts from any island domain consistently gives rise to worse effects than the displacement of a complement from the same domain:

29) i. *\"What, do you know [why Wally wore t₁]?\"
    ii. **\"Why, do you know [what Wally wore t₁]?\"

A similar extraction asymmetry was also noted between subject and objects:

30) i. *\"What, do you know [why Wally wore t₁]?\"
    ii. **\"Who, do you know [why t₁ wore a fez]?\"

Assuming that relative clauses are adjoined to the NPs that they modify, they too will not be properly governed since the lexical N° head will not govern an adjoined position. Thus the relative clause island can also be explained by the CED.
Huang (1982) is among the first to offer a systematic account of these extraction asymmetries, invoking the Empty Category Principle of Chomsky (1981), stated in (31):

31) The Empty Category Principle (ECP)
Every empty category must be properly governed

Informally, the ECP states that an empty category such as a trace is only licensed by satisfying proper government, as defined in (28) above. Thus a trace will have to be governed either by a lexical head or a co-indexed antecedent.

Originally, the ECP was devised specifically to account for so-called ‘that-trace’ phenomena described in section 1.2.6 and 1.3.3 of the previous chapter.

32) i. Who do you think [\(t_1\) found fruit on his doorstep]?
   ii. *Who do you think [that \(t_1\) found fruit on his doorstep]?
   iii. What do you think [that Mr. Nyman found \(t_1\) on his doorstep]?

Chomsky (1981) uses the ECP to explain the contrast between (32i) and (ii); the subject trace of “who” is said to be properly governed in the former but not the latter. In neither (32i) nor (ii) will the subject trace be governed by a lexical head, since subjects are specifiers of \(I^0\), a functional head, and thus to comply with the ECP they must be governed by a co-indexed antecedent. Assuming the existence of an intermediate trace in the embedded [spec,CP] position, in both cases this intermediate trace will be the closest co-indexed antecedent for the lower trace.

33) i. Who do you think [{\(CP\)} \(t'_1\) [\(IP\) \(t_1\) found fruit on his doorstep]]?
   ii. *Who do you think [{\(CP\)} \(t'_1\) that [{\(IP\) \(t_1\) found fruit on his doorstep}]]?

According to Chomsky (1981), the presence of the overt complementiser “that” in the \(C^0\) position of (32ii)/(33ii) prevents the intermediate trace (\(t'_1\)) in [spec,CP] from governing the root trace (\(t_1\)) in [spec,IP]\(^14\). Thus the lower trace remains ungoverned and the structure violates the ECP, causing ungrammaticality. In (32i)/(33i), however, the C head, being empty, does not prevent government of \(t_1\) in [spec,IP] by \(t'_1\) in [spec,CP], and thus the structure satisfies the ECP. Note too that the root trace in (32iii), being a complement of the verb “found”, will be head-governed by this lexical category. In this way the lower trace will satisfy

\(^{14}\)The fact that the presence of the complementiser “that” in \(C^0\) prevents \(t'_1\) from governing \(t_1\) has to be stipulated. See Chomsky (1981, 1986-a), Lasnik and Saito (1984, 1992) and Rizzi (1990) for further discussion of this point.
clause of the ECP, and thus the presence of the complementiser “that”, which prevents its antecedent government, is irrelevant.

Huang (1982) exploits the ECP in (31) along with the disjoint nature of proper government to account for extraction asymmetries between subjects/adjuncts and objects. As should be clear by now, the first criterion of proper government (government by a lexical head) will only ever be fulfilled by complements of lexical heads, such as verbs and prepositions. Traces of other elements, such as subjects and adjuncts, which are not properly governed by a lexical head, will have to satisfy the second, antecedent-government clause of (28). In this way, apart from obeying Subjacency and the CED, each trace in a chain headed by an extracted subject or adjunct will also have to be governed by a co-indexed antecedent in order to avoid falling foul of the ECP. Consider now the partial representation of (29ii) in (34) below:

Here, quite apart from violating Subjacency by crossing the two IP bounding nodes, movement of the adjunct “why” also fails to satisfy the ECP. The adjunct trace $t_j$ adjoined to the lower IP evidently cannot be properly governed by a lexical category since the head of this IP projection is functional, and will not govern an adjoined position anyway. Neither, however, is this trace properly governed by any co-indexed antecedent; the only element co-indexed with the trace $t_j$ is “why,” which is too far away from $t_j$ to govern it. “Why” in the matrix [spec,CP] position will only govern other elements within the same CP projection in which it itself is housed, the C⁰ head and IP along with the head and specifier of IP. There is simply no way that the governing domain of “why” can ‘reach’ to the $t_j$ trace adjoined to the embedded IP.
According to Huang (1982) this simultaneous violation of the ECP and Subjacency would account for the increased ungrammaticality of the example in (29ii)/(34) over (29i), which only violates Subjacency; from the representation in (22) above it should be clear that the root trace of “what” in (29i) will be properly governed by the lexical head “wore” and will thus satisfy the ECP irrespective of whether or not it is governed by a co-indexed antecedent. Exactly the same contrast applies in the case of the subject/object asymmetry illustrated in (30) above; while (30i) violates Subjacency as just discussed, (30ii) also falls foul of the ECP since the extracted WH-subject in the matrix [spec, CP] position will be too far away to antecedent-govern its trace in the embedded [spec, IP] position.

35) **[cp WhOj do [ip you know [cp why [ip ti wore a fez]]]]

Here, just as in (34), the presence of a WH-word in the embedded [spec, CP] position prevents an intermediate trace from occurring there which, in a grammatical structure, would properly govern the lower trace:

36) [cp WhOj do [ip you know [cp ti [ip ti wore a fez]]]]

Furthermore, the intermediate trace in the embedded [spec, CP] position which governs ti is itself properly governed by the lexical head “know”, and thus also satisfies the ECP.

I will say more about the interaction of the ECP with other island constraints in the following section. The important point to note here, though, is that the same configuration which gives rise to a Subjacency violation in (34) and (35) will also automatically yield an additional ECP effect in chains where the root trace is not properly governed by a lexical head. Generally, if movement from position A to B across a structure X violates Subjacency (or, for that matter, the CED) then the presence of X will also prevent a B from governing A. Thus traces of subjects and adjuncts, which rely on antecedent government to satisfy the ECP, will inevitably fail to comply with the ECP when extraction occurs from within an island domain. In this way the extraction of subjects and adjuncts from islands is predicted to give rise to worse effects than the extraction of complements, whose traces satisfy the ECP through head government.

\[15^{\text{Recall that a head X will govern its complement YP as well as the specifier and head of YP.}}\]
2.4.4 - Barriers.

Huang's (1982) work represents a significant advance in the study of extraction phenomena, and certainly achieves some major theoretical and empirical improvements over the Standard Theory account. However, one weak point of Huang's approach is the somewhat inelegant distinction between the twin constraints of Subjacency and the CED. On the one hand Subjacency, which accounts for the complex NP and WH-islands, prevents extraction across two (or more) bounding nodes, these defined, somewhat stipulatively, in terms of their category, IP and NP. On the other hand, the CED, which accounts for the subject and adjunct islands, proscribes extraction from domains not on the basis of their category, but according to their configurational relationship with a lexical governor.

In his 1986-a monograph 'Barriers' Chomsky seeks to eliminate this troublesome distinction between the CED and Subjacency by reformulating constraints on movement in terms of constraints on government. Put simply, Chomsky's solution is to redefine the concept of 'bounding node' in Subjacency as a barrier for government. In this way whatever constitutes a barrier for government will also be a barrier for movement, and thus any category which prevents A from governing B will also block movement between A and B. One major implication of this, of course, is that each link in a syntactic chain will have to satisfy a government relation as well as a c-command condition. We can thus perhaps already begin to see how this new conception of Subjacency may interact plausibly with the ECP discussed in the previous section. However, the ideas raised in Chomsky's monograph are very complex, and what follows is only a simplified outline of his proposals.

In defining the concept of 'barrier' (both to movement and government) Chomsky (1986-a) utilises yet another structural relation which he refers to as L-marking:

37) L-Marking

A L-marks B if ...
   i) A is lexical, and
   ii) A governs B, and
   iii) A assigns a θ-role (semantic role) to B

Informally, then, L-marking can be interpreted as a more restricted version of government which will only hold between a head and its arguments. A maximal projection (XP) can be a barrier in two different ways, either inherently or by inheritance; any XP which is not L-marked is described as an inherent barrier, with the exception of IP, which can never be an inherent barrier. In this way any XP (except IP) which is not governed and assigned a θ-role by a lexical
head will constitute a barrier to movement and government for whatever it dominates. Note, though, that the exception of IP is an unattractive stipulation.

The other way in which a maximal projection can be a barrier is by inheritance; an XP can inherit 'barrierhood' by immediately dominating another barrier, a maximal projection which is not L-marked. In this way even IP, or an L-marked maximal projection, which are not barriers inherently, can inherit 'barrierhood' by dominating another barrier. According to Chomsky (1986-a) a single barrier will block government while also causing a 'weak' Subjacency violation for anything moved across it. A more serious Subjacency violation will arise from crossing two or more barriers, irrespective of whether these are defined as such inherently or by inheritance.

However, this more stringent locality condition on movement implied by Chomsky's reformulation of Subjacency would mean that even perfectly grammatical examples of clause-bound object extraction would cross two barriers, VP and IP, and would thus be predicted to be ungrammatical. To avoid problems of this sort Chomsky proposes that even simple clause-bound movement takes place in two separate stages, with an extractee first adjoining to VP before moving on to the [spec,CP] position. Consider, for example the grammatical sentence in (38) with its associated structure:

38) Who did Pulat kill?

[Diagram of the structure of sentence (38)]

Here rather than moving to the matrix [spec,CP] position in one go the extracted object "who" first moves to a position adjoined to VP leaving an intermediate trace there. Neither this nor the subsequent movement from the adjoined position to [spec,CP] crosses any barrier; the adjunction of the intermediate trace to the VP in (38) has the effect of splitting this constituent into two parts, creating an 'upper' and 'lower' VP. The initial movement of "who" to the
adjoined position only passes the lower segment of this VP, and thus avoids crossing the entire VP, which would otherwise constitute an inherent barrier since it is not L-marked. Similarly, movement from the VP-adjoined position to the matrix \([\text{spec,CP}]\) will only cross the upper segment of the VP projection, and will thus once again avoid crossing the VP barrier itself. Moreover, since the upper segment of the VP dominating the intermediate trace in (38) does not in itself constitute an inherent barrier for this trace, the IP immediately dominating the VP segment will not become a barrier by inheritance. Thus movement from the VP-adjoined position to the matrix \([\text{spec,CP}]\) will be unproblematic. Notice too that quite apart from avoiding a Subjacency violation, the presence of the adjoined intermediate trace in (38) will also serve to govern (properly) the lower trace \(t_i\), although this trace is already properly governed anyway by the lexical head "kill". Similarly this intermediate trace will itself be (properly) governed by the co-indexed WH-word "who" in \([\text{spec,CP}]\) since the only maximal projection intervening between these two elements is IP, which does not constitute a barrier.

Chomsky’s revised formulation of Subjacency in terms of barriers accounts for the subject, adjunct and relative clause islands in a similar way to Huang’s CED. Neither a subject nor an adjunct, whether of a verb or a noun, will be L-marked (just as they are not properly governed) and thus they will always constitute barriers for material which they dominate. Consider, for example, the adjunct island violation in (39ii) below with its associated representation:

39) i. Pulat married Stachia \([\text{before he killed his brother}]\).

   ii. *Whom did Pulat marry Stachia \([\text{before he killed _}]\)?
This structure contains two intermediate traces, one adjoined to the VP of the adjunct clause and one occupying the [spec,CP] position of the same embedded clause. The first two movements of “whom”, from the complement position of “killed” to the VP-adjoined position and from here to the embedded [spec,CP], are both unproblematic since they cross no barrier. However, movement from the [spec,CP] of the adjunct clause to the matrix [spec,CP] position crosses two barriers in the form of the embedded CP and the matrix IP; being an adjunct, the embedded CP in (39ii) is not L-marked and is thus an inherent barrier. In addition, the matrix IP, although not an inherent barrier, becomes a barrier by inheritance, since it immediately dominates the adjunct CP. By crossing two barriers in this way the extraction of “who” in (39ii) gives rise to a severe violation of Subjacency.

Notice, however, that in (39ii) the same CP and IP barriers causing a Subjacency violation in the movement between the embedded and matrix [spec,CP] positions will also naturally prevent “whom” from governing the second intermediate trace ti”. This trace, being an empty category, should be properly governed in order to satisfy the ECP, and thus the natural, though incorrect, implication of this is that (39ii) will violate the ECP as well Subjacency. As I have already noted, however, examples of complement extraction like (39ii) are actually supposed to satisfy the ECP, thus explaining their increased acceptability over examples such as (40ii) which do not:

40) i. Pulat married Stachia [before he killed his brother with a pencil]
   ii. **How did Pulat marry Stachia [before he killed his brother _ ]?

In order to avoid this problem Chomsky adopts the complex and somewhat stipulative proposals of Lasnik and Saito (1984) involving the deletion of intermediate argument traces before LF, where they assume that the ECP is checked. For our purposes the result of these proposals is that intermediate traces of extracted objects are, in effect, immune to effects of the ECP. Thus the structure in (39ii) will violate Subjacency but not the ECP since the un governed ti”” trace in the [spec,CP] position of the adjunct clause is simply ‘irrelevant’ to the ECP. If, however, an adjunct or subject were extracted from the adjunct CP in (39ii) then its un governed intermediate trace in the embedded [spec,CP] position would cause a violation of the ECP, since, according to Lasnik and Saito (1984) the intermediate traces of these elements are not immune to the ECP. Thus Chomsky (1986-a) follows Huang (1982) in attributing the asymmetry between the ungrammatical examples in (39ii) and (40ii) to the added effects of the ECP in the latter but not the former.
Chomsky's reformulation of Subjacency in terms of barriers also accounts for the subject island; (41ii) offers a relevant example:

41) i. [Eating bat curry] might cause heartburn.
   
   ii. *What might [eating _ ] cause heartburn?

Here the extracted word "what" reaches the specifier position of the subject CP via the VP-adjoined position without any difficulty. However, movement from the embedded [spec,CP] to the matrix [spec,CP] position crosses two barriers, the subject CP and the matrix IP. Firstly, even though the subject clause may be assigned a 0-role by "cause", it is evidently too far away to be governed by this verb, and will thus not be L-marked according to the definition in (37). Consequently the subject CP in (41ii) will constitute an inherent barrier to movement and government for the trace it dominates. Furthermore, the IP node, by immediately dominating this CP barrier, will itself become a barrier by inheritance. Thus movement from the [spec,CP] position of the embedded subject clause to the matrix [spec,CP] position crosses two barriers, causing once again a severe violation of Subjacency.

Once again, the CP and IP barriers between "what_" and the intermediate trace \( t_i' \) in (41ii) will prevent the former from governing the latter. As with the example in (39ii), this will not cause a violation of the ECP here since, according to the assumptions of Lasnik and Saito (1984), intermediate traces of extracted complements are immune to its effects. Nevertheless, extraction of a subject or adjunct from the subject clause in (41ii) would leave an intermediate trace in the [spec,CP] position of this clause which did violate the ECP in not being properly governed, either by a lexical head or by the co-indexed antecedent "what_" in the matrix.
[spec,CP] position. Once again, then, the added effects of the ECP are taken to explain the increased ungrammaticality of examples like (42ii) below, where an adjunct is extracted from the island domain:

42)  i. [Eating bat curry at night] might cause heartburn.
    ii. **When might [eating bat curry _ ] cause heartburn?

As well as subsuming the subject and adjunct islands, Chomsky’s reformulation of Subjacency also accounts for the WH and complex NP islands. (43) below offers an example of the latter constraint:

43)  i. I heard [rumours that they arrested Balanescu].
    ii. *Who did you hear [rumours that they arrested _ ]?

Here the extractee “who” will reach the [spec,CP] position of the embedded clause without difficulty. However, movement from here to the VP-adjoined position in the matrix clause crosses two barriers, CP and NP; assuming that the noun “rumours” will not assign a θ-role to its complement CP, this CP will not be L-marked and will thus constitute an inherent barrier
for the $t_1'$ trace it dominates. Moreover, the NP node dominating this CP will inherit barrierhood from it, even though it is itself L-marked by the verb "hear". In this way movement from the embedded [spec,CP] to the VP-adjoined position of the matrix clause will violate Subjacency by crossing two barriers, explaining the ungrammaticality of the sentence in (43ii). Extraction asymmetries between objects and subjects/adjuncts will be explained by reference to the ECP in exactly the same way as with the examples in (39ii) and (41ii)\(^\text{16}\).

Through his reformulation of Subjacency in terms of barriers Chomsky (1986-a) is able to offer a fairly uniform account of virtually all the island data explored in the previous chapter. More specifically, with the replacement of the categorially-defined concept of bounding node by the (largely) configurationally-defined notion of barrier, it is possible to eliminate Huang's (1982) troublesome distinction between Subjacency and the CED. Very simply an 'island' domain preventing extraction will now be defined as any phrase (XP) which is not L-marked, this serving as a barrier to movement or government for anything it dominates. However, in spite of the conceptual advantages of Chomsky's (1986-a) account of island constraints, certain problems remain. For one thing, the stipulation that IP cannot be an inherent barrier is unattractive, and is actually no less arbitrary than stating that IP is a bounding node, as was the case in older versions of Subjacency. Furthermore, in following Huang (1982) and attributing extraction asymmetries between objects and subjects/adjuncts to the effects of the ECP, Chomsky is forced to adopt the complex and stipulative proposals of Lasnik and Saito (1984) concerning the intermediate traces of objects.

There are a number of other problems associated with the Barriers framework, some of which I will discuss further in section 2.5. However, it is important to bear in mind that Chomsky does not present his (1986-a) monograph as anything other than an exploratory work, and does not pretend to offer a fully-developed account of extraction. There have since been a number of attempts to improve on the 'barriers' framework described here. Notable among these are Rizzi's (1990) theory of Relativised Minimality and Manzini's (1992) Locality Theory. Both of these frameworks modify Chomsky's (1986-a) proposals while at the same time trying to address some of its weaknesses which I outlined above. However, I do not have space to review these accounts here.

\(^{16}\text{However, I should point out that this account of the complex NP island relies crucially on the (somewhat tenuous) assumption that the noun "rumours" does not assign a $\emptyset$-role to, and thus does not L-mark, its complement clause (Chomsky 1986-a p. 35).}\n
68
2.4.5 - Parasitic Gaps

In sections 1.2.7 and 1.3.4 of the first chapter I referred to the phenomenon of parasitic gaps (PGs), illustrated in (44ii):

44) i. Mike took a mango without paying for it.
   ii. What did Mike take _ without paying for _?

Here the extractee “what” is associated with two ‘gaps’, the second of which, in the adjunct clause, is said to be parasitic on the first. Cases of multiple gapping such as these pose a significant problem for any theory of extraction, since they appear to contravene the otherwise valid generalisation that each extractee is uniquely associated with a single ‘gap’ in the subsequent clause; after all, a single word cannot be extracted from more than one place.

Initially, it might appear that PP Theory has a ready-made solution to the problem of PGs in the form of intermediate traces. Thus the structure in (44ii) would be analysed as in (45), where the ‘gap’ following “take” is in fact occupied by an intermediate trace of “what” on its way from the complement position of “(paying) for” to the matrix [spec,CP]:

45) What; did Mike take t_1 without paying for t_2?

However, the analysis in (45) is in fact impossible. For one thing, irrespective of whether the adjunct clause is adjoined to the matrix IP or VP, the complement position of “(paying) for” will be higher up in the tree structure than the complement position of “take”. In this way the putative intermediate trace t_1 will fail to govern (or even c-command) the root trace t_2, which is predicted to result in a Subjacency violation. Moreover, the analysis in (45) is also ruled out by a general constraint preventing chains, as well as nominals, from being assigned more than one θ-role (Chomsky 1986-b, Brody 1993, 1995). In (45) the chain <What, t_1, t_2> would be assigned two θ-roles, one from “take” and one from “(paying) for”.

In response to these problems Chomsky (1986-a) is forced to conclude that structures like (44ii) in fact display two separate instances of extraction. According to this analysis the gap in the matrix clause will arise normally from the movement of “what” to the [spec,CP] position:

46) What; did Mike take t_1?

However, the second, ‘parasitic’ gap in the adjunct clause is said to arise from the movement of a non-overt empty operator (O) to the [spec,CP] position of this adjunct clause:
47) What did Mike take \( t_i \) \([\text{CP}\ O_j \text{ without paying for } t_j]\)?

Here then we have two separate chains, \(<\text{What}_i \ldots t_i>\) and \(<O_j \ldots t_j>\), each of which is assigned its own distinct \(\theta\)-role. According to Chomsky (1986-a p. 56) these two chains could themselves constitute a "composed chain" \(<C_1 \ldots C_2>\) where \(C_1 = \langle\text{What}_i \ldots t_i>\) and \(C_2 = \langle O_j \ldots t_j>\). Within this composed chain the presence of \(C_2\) would be conditional upon the presence of \(C_1\), thus accounting for the fact that the gap associated with the moved operator is dependent, or 'parasitic', on the first gap arising from the extraction of "what".

One of the advantages of the analysis in (47) is that it explains the fact, noted in section 1.3.4, that the PG in the adjunct clause does not give rise to a violation of the adjunct island; the non-overt operator \(O_j\) is not actually extracted from the adjunct clause but is instead moved to a different position within the same clause. However, as far as I know, this analysis offers no explanation for the fact that extracted adjuncts do not license PGs. There is, in other words, nothing to prevent an ungrammatical structure such as (48ii):

48) i. *Why did Mike take a mango \_[ without paying for it \_]?
   ii. *Why \_ did Mike take a mango \_ [Oj without paying for it \_]?

In section 2.3.3 I described how in PP Theory a word can move covertly, between S-Structure and LF. Now, though, the analysis of PGs in (47) also allows for the 'overt' movement (between D-Structure and S-Structure) of a non-overt category. 'Invisible' derivations such as these, with no reflex in the overt syntax, add greatly to the power of a theory, and, unless very carefully controlled, could be argued to increase its potential to overgenerate.

2.5 - Appraisal of the 'Barriers' Account

My aim in this chapter has been to outline some of the many strands of the Principles and Parameters (PP) Theory account of extraction phenomena. In particular I hope to have drawn attention to how this broad approach is characterised by two fundamental components of the theory, phrase structure and syntactic movement, the latter being part of PP Theory's transformational heritage. These two specific components of PP Theory are of particular relevance since in many ways their absence is a defining characteristic of the dependency-based accounts of extraction that will be discussed later. Very simply, PP Theory conceives of

\[\text{For the sake of simplicity I have omitted any intermediate VP-joined traces from these chains.}\]
extraction as syntactic movement of a constituent from one configurationally-defined position to another. This movement, which may occur in two or more separate stages, is constrained by a range of structural relationships which must obtain between the extractee itself and its trace(s). Failure of these relationships to apply will result in an island violation.

In section 1.4 of the first chapter I suggested that an account of extraction should be both holistic, in accounting for as wide a range of data as possible, and realistic, in accounting for these data by reference to as small and constrained a syntax as possible. The main reason for including the discussion of PP Theory in this chapter is that it has made more progress than any other model of syntax in seeking to attain these goals. As far as I am aware, no other framework is able to account for such a broad range of extraction data, nor has been as successful in attempting to reduce these complex data to a limited set of underlying syntactic principles. For these reasons it is fair to say that the PP Theory account of extraction sets the standards against which other theories are judged, and consequently the discussion in this chapter will provide an essential context for the exploration and evaluation of the dependency-based approaches to extraction in the following chapters.

However, the relative success of PP Theory in accounting for extraction data does not mean that this account is without its own weaknesses and problems. I have already pointed out in section 2.4.4 some of the stipulations within the Barriers framework, such as IP’s exceptional status as a non-barrier. Furthermore, some of the complex structural relations which play such a key role in this framework may also appear to be somewhat arbitrary in nature. For example, while Chomsky (1986-a) defines the notion of barrier in terms of L-marking, no explanation is offered as to why the three criteria inherent in the definition of L-marking in (37) should be relevant to ‘barrierhood’. After all, to state that extraction is only possible from a domain which is governed and assigned a θ-role by a lexical head is no more explanatory than stating that extraction is only possible from the complement of a verb or preposition, a generalisation which also ‘accounts’ for a variety of island phenomena. It seems equally arbitrary to state, as in (28), that only a lexical head or co-indexed category should constitute a ‘proper governor’.

There is undoubtedly more to be said about the formulation of some of these complex relations and their role in PP Theory. Of course some people go still further and reject the very concepts upon which the theory is based, such as the notions of syntactic movement and empty category. I do not have time to explore these issues further here, though. As far as its ‘holism’,
or observational adequacy, is concerned, PP theory is undoubtedly successful in accounting for the majority of the data discussed in chapter 1. Chomsky’s (1986-a) framework captures virtually all of the island constraints, with the curious exception of the co-ordinate island which, since the work of Ross (1967), has remained largely ignored within PP Theory generally. In addition Chomsky (1986-a p. 35) notes a possible problem his theory faces with the complex NP island.

However, as it stands the Barriers framework also encounters a problem in connection with the subject island. As I noted in section 2.2.1, subjects in PP Theory, like objects, are defined according to their position in the phrase structure, usually identified as the specifier position of IP. In this way, according to Chomsky (1986-a) as well as later approaches (Manzini 1992), subject clauses are always predicted to be islands since they will never be L-marked. As I noted in section 1.3.2 of the first chapter, though, the status of subject clauses as islands seems to have more to do with their position relative to the matrix verb rather than any inherent structural property. Thus I pointed out cases in English and Modern Greek where extraction is equally acceptable from subject and object clauses when they follow the matrix verb, and equally unacceptable when the same clauses precede the matrix verb. I suggested that a ban on extraction from subordinate clauses which precede the matrix verb might also explain the absence of extraction in SOV languages, as well as the absence of subject island effects in at least some VSO languages (Chung 1983). PP Theory would have some difficulty in accounting for these data given that subjects will still occupy the [spec,IP] position, and thus block extraction, irrespective of their actual linear placement in a structure.

Related to this last point is the question of extraction asymmetries between subjects and objects. In section 2.4.4 I described how Chomsky (1986-a), following Huang (1982), invokes the ECP to explain this contrast. Once again, then, the fact that subjects are less extractable from island domains than objects is linked to their inherent structural property of not being governed by a lexical head. The implication of this of course is that the extraction of a subject from an island will always violate the ECP and will thus inevitably be more problematic than the extraction of an object. As I pointed out in section 1.3.3 of the previous chapter, however, this is not always true, and it is possible to find examples where subjects behave more like objects than adjuncts with respect to their extractability. These cases include ‘weak’ islands, which pose no significant problem for the displacement of subjects or objects, although they do for adjuncts:
49) i. They don't think [Ted crashed his Bentley because he was drunk].
   ii. What don't they think [Ted crashed _ because he was drunk]?
   iii. Who don't they think [_ crashed his Bentley because he was drunk]?
   iv. Why don't they think [Ted crashed his Bentley _ ]?

Similarly, subjects seem to pattern with objects rather than adjuncts when extracted from a position embedded within an island:

50) i. I know [why Rudolf thinks Santa stole the cash].
   ii. What do you know [why Rudolf thinks Santa stole _ ]?
   iii. Who do you know [why Rudolf thinks _ stole the cash]?
   iv. Who do you know [why _ thinks that Santa stole the cash]?

As I noted in section 1.3.3, the examples in (50ii) and (iii) strike me as being equally bad, and considerably better than (50iv). However, according to the PP Theory account both (49iii) and (50iii) would violate the ECP, incorrectly predicting a degree of ungrammaticality comparable to (49iv) and (50iv) respectively. In this way, by attributing genuine asymmetries between subjects and objects to inherent structural distinctions between the two, PP Theory will inevitably fail to account for those cases where these asymmetries do not apply.

One weakness which PP Theory shares with most, if not all, syntactic accounts of extraction is its inflexibility. Syntactic theories generally maintain an absolute distinction between grammaticality and ungrammaticality. However, I believe that this sort of rigid binary approach is inappropriate for the notable scalability of ungrammaticality associated with island violations. It is true that PP Theory is better than some in this respect in that it at least allows the effects of the ECP to override Subjacency violations, making some examples worse than others. However, the examples in (51) attest to the range of ungrammaticality associated with different islands, even when the same constituent is extracted:

51) i. What did they make [a claim that Sid ate _ ]?
   ii. What did they discuss [a claim that Sid ate _ ]?
   iii. What did Sid smirk [while eating _ ]?
   iv. What did they ask [why Sid ate _ ]?
   v. What did they know [a man who ate _ ]?

Since an object is extracted from each of the islands here, the effects of the ECP will be irrelevant. According to the 'Barriers' framework discussed in section 2.4.4, the movement of
“what” in each of these examples will cross two barriers, resulting in a severe violation of Subjacency. In this way the five sentences in (51) are predicted, incorrectly, to be equally bad.

Chomsky (1986-a) does suggest that milder violations of Subjacency might arise from crossing a single barrier. However, within his framework this only applies to the WH-island, accounting for the relative acceptability of examples like (52):

52) *Whom did she ask [where to meet _ ]?

However, this still fails to explain the contrast between (52) and its tensed equivalent in (53):

53) *Whom did she ask [where they met _ ]?

In response to this point, Chomsky (1986-a p. 37-38) suggests that a tensed IP in subordinate clauses might, after all, constitute an inherent barrier. However, this solution, apart from being stipulative, only applies in explaining the contrast between a ‘one-barrier’ violation such as (52) and a ‘two-barrier’ violation like (53). In section 1.3.1 of the previous chapter, though, I described how tense actually plays an important role in determining the severity of all island violations. As it stands the ‘Barriers’ model would be unable to account for the majority of these tense-related contrasts.

In spite of my attempts to pick holes in it here, there is no doubt that the PP Theory account of extraction phenomena remains the most comprehensive and thorough treatment of these data, and the one which sets the standards which other theories must try to match or, possibly, surpass. In recent years, however, with the development of the Minimalist Programme, extraction-related issues have not received the same attention as they once did, and it is too early to say whether the complex account of these data developed in the PP framework can be carried forward into the theory’s new stage of evolution.
CHAPTER 3

DEPENDENCY, WORD GRAMMAR AND EXTRACTION

3.1 - Introduction

In the previous chapter I outlined some key points of Principles and Parameters (PP) Theory and some of the various accounts of extraction phenomena proposed within this broad framework. In this chapter I will turn my attention to a very different theory of language, Word Grammar (WG), a dependency theory, along with its own account of the same data. As I pointed out in the first chapter, extraction has not received the same attention within dependency grammars as it has in the constituency-based linguistic mainstream, and the central goal of this thesis is to explore further the potential of a dependency-based theory of syntax to account for the complex displacement data discussed so far. My reasons for selecting WG as representative of the dependency grammar tradition are twofold: Firstly WG is the best-developed contemporary dependency theory, and the only one with a relatively abundant and easily-accessible source of literature. Secondly, as I pointed out in section 1.4 of the first chapter, WG is the only true dependency theory to date in which a genuine attempt has been made to offer a serious account of extraction data.

The examination of WG within the context of PP Theory is meant to be more than a matter of simple theory comparison; in order to understand fully a theory’s approach to a specific range of data (such as extraction) it is, I believe, crucial to understand at least something about how that approach is determined and shaped by wider, more general assumptions. Moreover, issues arising from the comparing and contrasting of PP Theory and WG will provide a useful context for the exploration of an alternative approach to dependency grammar and extraction phenomena that will be discussed in the two subsequent chapters. Section 3.2 of this chapter outlines some of the basic characteristics of dependency grammar and explores the key differences between a dependency-based and constituency-based formalism. Section 3.3 deals with some of the broader issues concerning WG’s conception of linguistic knowledge and grammar, while sections 3.4 and 3.5 offer an outline and evaluation of WG’s account of extraction.
3.2 - An Outline of Word Grammar Syntax

3.2.1 - Dependency Syntax

One way of expressing a syntactic relationship between two words is by means of constituency. In its simplest form this view of syntax allows two words, A and B, to be linked by their shared participation in a phrasal constituent:

1) \[ \begin{array}{c}
A \\
\text{AP} \\
\text{BP} \\
B
\end{array} \]

These phrases can then be embedded within one another in various ways, giving rise to the type of branching configurational structure discussed in chapter 2.

In the previous chapter I described how endocentric constituent structure of the type illustrated in (1) was an integral part of the linguistic mainstream, exemplified by PP Theory and its recent Minimalist extensions. Much of the machinery of these theories is geared around the putative constituency obtaining between any two words which enter into a syntactic relationship. Indeed, similar configurational patterns are even taken to extend beyond syntactic relations; not only does one encounter phrases headed by 'functional' elements such as C and I, one also finds a similar pattern generalised to the internal morphological structure of words (Williams 1981, Selkirk 1982). According to this constituency-based approach to language, then, different syntactic relations between words will be defined and distinguished on the basis of geometrical properties of phrase structure; recall, for example, that in PP Theory relations such as subject, object and adjunct have no primitive status, and are all defined in geometrical terms such as dominance, sisterhood and bar level.

However, as I noted in section 1.4 of the first chapter, phrase structure is not the only way in which words can be brought together. A possibly simpler approach might be to take the syntactic relationships which link words as basic, and not as being derived from or mediated by abstract configurational structure. In this way word A and word B will be linked directly by a simple relation, R, without participating in any form of higher phrasal constituent. Any connection between A and B will be expressed solely in terms of R:

2) \[ \begin{array}{c}
A \\
R \\
B
\end{array} \]

This, in a nutshell, is the view of syntax enshrined in Dependency Grammar (DG), a system of linguistic analysis which goes back to ancient times, but which finds its modern theoretical
roots in the work of Tesnière (1959). The best-developed and most fully-articulated contemporary theory of DG is Word Grammar (WG), developed over the last twelve years or so by Richard Hudson (1984-a, 1987, 1988-a, 1988-b, 1990, 1994). Other contemporary dependency-based theories of syntax include Lexicase (Starosta 1988) and Mel’čuk’s (1988) unnamed theory; Fraser (1993) and Hudson (1993) give a more detailed survey of the current trends in DG. In addition, Relational Grammar (Perlmutter 1983, Perlmutter and Rosen 1984, Blake 1990), Case Grammar (Fillmore 1968, Anderson 1971, 1977) and Categorial Grammar (Ades and Steedman 1982, Oehrle et al. 1988, Wood 1993) are also sometimes described as dependency grammars, in that they treat syntax in terms of direct relationships between words rather than constituency. These theories, however, while retaining some close affinities with DG proper, embody important differing assumptions and formalisation, and are perhaps best-described as first cousins of DG.

In common with other theories of dependency syntax, WG eschews constituency of all types above the level of the word, and syntactic representations consist solely of word strings with individual pairs of words linked together by asymmetric binary relations called dependencies. (3) below shows an example of a structure, with each dependency relation represented by an arrow:

3) 
Santa stole the cash yesterday.

According to this view, then, instead of analysing a sentence by breaking it up into successively smaller units until we reach word-level, instead we take words themselves as our starting point and derive phrases and sentences on the basis of how words are linked with one another, as Mel’čuk (1988 p. 4) concludes: “constituency is a MANIFESTATION of syntactic structure, not syntactic structure itself” (author’s italics). In this way the structure of a sentence is basically a reflex of how its component words relate to one another. Words thus assume a central importance in WG, a fact from which the theory derives its name (Hudson 1984-a).

1Lexical Functional Grammar (Bresnan 1982) is exceptional in that it recognises both phrase structure and dependencies as primitive components of syntax.

2However, an exception is made for co-ordination structures (Tesnière 1959, Hudson 1988-a).
The arrows in (3) serve to express both the location and the asymmetry of dependency relations. Thus we see a direct syntactic relation between “Santa” and “stole”, for example, but not between “Santa” and “cash” or “yesterday”. The element of asymmetry or inequality is a fundamental characteristic common to all conceptions of dependency, both linguistic and non-linguistic; beyond the realm of language, we might describe Gibraltar as a dependency of the United Kingdom, but not vice versa. Dependency, no less in its linguistic sense than in its more general usage, incorporates the idea of one element in the relationship being somehow ‘more important’ and governing or controlling the other. Within a syntactic dependency relation this asymmetry is instantiated between a head word and a dependent word. The head is described as the governing element to which the dependent is subordinated, in a sense to be discussed later. Arrows point uniformly from heads to their dependents; thus “Santa”, “the” and “yesterday” in (3) are each dependents of the verb “stole”.

3.2.2 - Dependency and Constituency

One question that naturally arises is to what extent dependency and constituency are substantially different. Does the inclusion of A and B within a constituent AP or BP amount to the same thing as their participation in a relation R? What in other words is the real difference between (4i) and (4ii) below?

4) i. \[ \text{AP} \quad \text{B} \] ii. \[ \text{A} \quad \text{R} \quad \text{B} \]

Certainly (4i) and (4ii) serve to express similar information, and in some sense could be said to be equivalent (Hays 1964, Gaifman 1965, Robinson 1970, Hudson 1984-a, 1995-b, forthcoming-a); asymmetric dependency relations between heads and dependents can certainly be discerned in phrase structure where, in simple terms, a word A constitutes the head of its projection (AP). With respect to (5) below, BP and CP could be described as dependents of A; in some sense both BP and CP are subordinated to A by virtue of the fact that they occur embedded within A’s projection:

5) \[ \text{AP} \quad \text{BP} \quad \text{A} \quad \text{CP} \]

Note, though, that asymmetry is here derived from geometrical attributes of the phrase
structure configuration whereas in Dependency terms it is stated as a primitive property of syntactic relations. Provisionally, then, we might say that (5) is equivalent to (6):

6) \[ \text{BP} \rightarrow A \rightarrow \text{CP} \]

Bearing in mind, though, that the projections BP and CP in (6) really serve only to capture B and C’s respective dependency structure, a more accurate equivalent of (5) will be (7):

7) \[ \text{B} \rightarrow A \rightarrow \text{C} \]

In fact any endocentric constituent structure will have a ‘flat’, dependency-type analogue (Matthews 1981 ch. 4). In spite of this partial equivalence of constituency and dependency, however, there remain important differences between the two systems; as Buszkowski (1988 p. 92) (quoted in Wood 1993) notes, “... that two kinds of grammar are weakly equivalent does not mean, obviously, that they are simply the same”. Below are listed five points in which constituency and dependency systems differ.

i - It has been claimed that a dependency structure like (7) is simpler and more economical than the equivalent constituent structure in (5) since the latter incorporates non-terminal nodes, such as A’ and AP, which are absent in dependency structure (Hudson 1984-a). While this is true, it is only fair to point out that the primitive dependency relations in (7) are absent in (5). There is, however, a more indirect sense in which a dependency structure might be more simple than at least certain versions of constituent structure. In the previous chapter I drew attention to the fact that in PP Theory, the primary relations of phrase structure such as dominance and sisterhood must be supplemented with a system of secondary relations such as c-command and government. A dependency-based theory which can do without any such system of derived syntactic relations will be more economical than a constituency-based theory which cannot.

ii - Another important difference between a constituent like (4i) and a relation such as (4ii) is that only the former may itself participate in a relation with another word; whereas a constituent, such as CP in (5), can combine with something else (A) to make another constituent (A’), a relation (R) between two words cannot itself enter into another relationship with a third word. Only the words which participate in R can themselves relate to other words outside R. This means that from an external point of view, a relation, unlike a constituent, will
always be subordinate to its two component words. As Hudson (1990, 1995-b) notes, this can have some undesirable consequences for phrase structure theories, notably in cases where a direct relation between words is desirable but impossible due to the presence of an intervening phrase. Hudson (1995-b) cites the example of the verb “depend”, which requires a specific preposition, “on”, as a complement. In a constituency-based system the PP node intervening between the verb and the preposition itself would make any direct link between the two very difficult to express.

iii - While it is generally true that any well-formed phrase structure representation will have a dependency-type analogue, the converse is not always true; there are certain structural permutations which are inherently possible in dependency but which are excluded by very general properties of phrase structure. These permutations include mutually-headed and multi-headed constructions, (8) and (9), as well as structures involving crossing dependencies (10):

8) 
\[ A \rightarrow B \]

9) 
\[ A \rightarrow B \rightarrow C \]

10) 
\[ A \rightarrow B \rightarrow C \rightarrow D \]

All of the above permutations would be excluded by general properties of constituent structure such as modifier maximality (Jackendoff 1977), feature permeation (Muysken and Riemsdijk 1986) and binary branching (Kayne 1984, Larson 1988). Of course if structures corresponding to those in (8)-(10) were shown to be absent in language, then they would have to be excluded from a dependency system by the imposition of separate principles. For example most theories of dependency incorporate a locality constraint, to be discussed in the next section, preventing dependencies from linking two words which are insufficiently local, ruling out (10). Similarly most dependency grammars disallow multi-headed and mutually-headed constructions like (8) and (9) (Gaifman 1965, Robinson 1970, Mel’čuk 1988). WG, however, is exceptional in this respect in that it allows structures of these types (Hudson 1984-a, 1988-b, 1995-a). Indeed, as I will show, the theory’s account of extraction phenomena relies crucially on the fact that a displaced element is simultaneously the dependent of more than one head.
iv - The phrase structure configuration described above is composed of uniformly structured constituents recursively embedded inside one another. It follows that the number of Grammatical Relations (GRs) which can be derived on the basis of this structure will be strictly limited. In fact, as I pointed out in the previous chapter, only three GR’s are generally derived from geometrical properties of phrase structure - subject (daughter of XP sister of X’), object (daughter of X’ and sister of X) and adjunct (adjoined to XP). In Dependency Theory however, GRs like subject and object are recognised as primitive entities of the grammar, each GR constituting a specific instantiation of the more general dependency relation (see section 3.2.4 below). For this reason DG effectively imposes no inherent constraint on the number and nature of these relations which may be recognised. In effect, then, R in (4ii) above could construed as a variable over any possible relation. As I will go on to discuss in chapter 4, this can have certain undesirable consequences for a dependency theory.

v - Just as a phrase structure configuration strictly limits the number of GRs that can be derived on the basis of it, so too similar constraints apply to how many of these derived relations may exist within a particular structure. As I described in Chapter 2, apart from adjuncts, which can occur recursively, general principles of X-bar theory ensure that each head can only have one complement and one specifier position. This means that a verb may only ever have one subject and one object. While this is potentially desirable for reasons of constrainedness, it does raise problems in certain circumstances, notably when a verb has more than one complement, such as a direct and indirect object for example. In these cases a strictly binary branching phrase structure must resort to ad hoc devices such as recursive VPs (Larson 1988, 1990). This, of course, raises separate questions for a phrase structure theory; what, for example, prevents us from having multiple IPs and thus allowing more than one subject?

The differences between constituency and dependency is too broad an issue to be examined adequately here, and the five points above are far from exhaustive (Hudson 1984-a, Mel’čuk 1988). Nevertheless from the above we can see that the equivalence between constituency and dependency masks a series of real differences between the two. More importantly, however, of the five points listed above, the last three together express a greater degree of freedom and flexibility inherent in a dependency system which distinguishes it from phrase structure - at least the relatively constrained and rigorous version of phrase structure adopted within PP Theory following the work of Chomsky (1970), Jackendoff (1977) and Kayne (1984, 1994). This added freedom and flexibility of dependency is of particular
relevance because, as I indicated in the second chapter, it was the perceived inflexibility of phrase structure, particularly its apparent inability to cope with anything other than strictly local relations, that necessitated the introduction of the transformational component in the Standard Theory. As I will show later, WG exploits this increased flexibility inherent in a dependency formalism to formulate a monostratal, non-transformational account of extraction.

3.2.3 - Dependency Structure

With the exception of co-ordination structures, WG does not recognise any grammatical entity larger than the word, and consequently all syntactic knowledge is expressed exclusively in terms of words and the dependencies which serve to link them. Generally, a grammatical (and unambiguous) sentence will be uniquely associated with a single well-formed network of dependency relations between words; a network of dependencies is judged to be well-formed if it obeys certain constraints, two of the most important of which are known colloquially as ‘no tangling’ and ‘no dangling’ (Hudson 1994). The former constraint, also known as projectivity, disallows relations to intersect one another, thus imposing a degree of locality on dependencies. This serves to rule out examples such as (11) below:

11)

*Santa stole the yesterday cash.

Here the relation between “the” and “cash” is insufficiently local in that another word, “yesterday”, intervenes which bears no relation to either of them. Note, though, following the discussion in the previous section, that this well-formedness constraint is not inherent to dependency, and must be stated as a separate principle. In Chapter 5, though, I will suggest that a similar locality principle can in fact be derived from processing factors.

The second key constraint on dependency representations, ‘no dangling’, simply requires that every word in a single well-formed structure be linked to at least one other word; no word can remain unconnected to any of the others, hence the ungrammaticality of (12) where the word “lychee” could be said to ‘dangle’:

12)

*Santa stole the cash yesterday lychee.
The ‘no-dangling’ constraint guarantees that all words in a structure will depend on another word in the same structure. The only exception to this is the root word where chains of dependency originate and which itself depends on nothing. Each structure will have one root word which has so far been distinguished by the absence of any arrow pointing to it. In WG, however, the root word is usually identified by a unconnected downward-pointing arrow. In most cases the root word will be the matrix finite verb, as with “stole” in (13) below:

13)  
Santa → stole → the → cash → yesterday

Generally, then, the root word of a sentence will be the matrix verb, and all other words within the same sentence will depend on it, either directly or indirectly. For example arguments and adjuncts of the matrix verb will all be dependents, as is clear from (13). This is no less true if one of these arguments or adjuncts is itself a verb:

14)  
Rudolf → said → Santa → stole → the → cash

Here the verb “stole” which has its own subject and object dependents, is itself the dependent of the matrix verb “said”. Note that in (13) and (14) the noun “cash” is a dependent of the determiner “the”. In WG all determiners, including demonstratives and possessive pronouns, function as heads of their nouns, and Hudson (1984-a), was actually the first to advance this proposal, subsequently made famous by Abney (1987) and adopted in PP Theory (see also Horrocks and Stavrou (1987)).

Facts pertaining to word order in WG are expressed by a head parameter, which states a generalised ordering between heads and their dependents. Thus in some ‘head-first’ languages, like Welsh and Somali, heads nearly always precede their dependents, while ‘head-last’ languages, like Turkish and Japanese, display a dependent-head ordering. The generality of the head parameter has the potential to subsume a wide range of word order phenomena, such as the ordering between complementisers and verbs, verbs and their arguments, prepositions/postpositions and nouns as well as determiners and nouns (Greenberg 1966, Comrie 1981). Of course, some languages are better-behaved than others with respect to the head parameter. Thus in English, although heads generally precede their dependents, certain dependencies do not conform to this pattern, the most common exception being subjects which
invariably precede their head.\(^3\)

WG is a monostratal theory in which each (unambiguous) sentence is uniquely associated with a single syntactic structure. Thus WG has no equivalent of Deep Structure, Logical Form or the transformational mechanisms which were shown to characterise PP Theory. Consequently empty trace elements resulting from movement are also absent in WG. It isn’t that analyses involving transformations and empty categories are necessarily incompatible with a dependency framework, but rather these components are rendered somewhat superfluous by the increased flexibility of dependency representations over constituent structure. As I suggested in section 3.2.2, all syntactic relations can be expressed by direct dependencies, irrespective of whether or not they are strictly local. Thus, theoretically at least, a displaced WH-object could depend on a head verb in exactly the same way as a non-displaced, non-WH equivalent.

3.2.4 - Dependencies and GRs

In example (13) above the verb “stole” is shown to have three dependents, which we might informally label as subject - “Santa”, object - “the cash” and adjunct - “yesterday”. In WG these Grammatical Relations, or GRs, are recognised as primary syntactic units of the theory, and thus each dependency relation must be labelled according to which of these more specific GRs it instantiates (s = subject, o = object, a = adjunct, c = complement):

15)

```
Santa stole the cash yesterday.
```

Each GR is described as a distinct subtype of the dependency relation, and as such, each is associated with its own particular set of properties. Thus GRs may differ from one another with respect to their distribution, direction, morphological marking or associated semantic content, amongst other things. For example, while certain verbs may have a single object dependent, which will usually occur immediately after its head, any verb can potentially have any number of adjuncts, the position of which is somewhat less restricted. The question of how

\(^3\)Adjectives too are exceptional in that they are described as dependents of the nouns which they precede (Siewierska 1988, Hudson 1990). In Kreps (1996-a), however, I argue that adjectives, like determiners, could instead be analysed as heads of the nouns they modify. One of the many advantages of this proposal, of course, is that it reduces the number of exceptions to the general head-dependent ordering pattern in English.
many separate GRs need to be recognised in an adequate theory of grammar is a controversial one. Most relational theories recognise a class of GRs which includes at least subject, object and indirect object along with various types of adjunct. These relations will be familiar since they play a part, at least informally, in most other theories.

In addition to these fairly well-known GRs, WG also recognises a number of more theory-particular relations including, for example, 'extraposee' and 'visitor', instantiations of which mediate between a displaced dependent and its non-local head (Hudson 1988-b). The visitor relation will be the subject of later sections of this chapter. Another GR particular to WG (and LFG) is the x-complement. Informally, x-complements could be described as complement verbs which share their subject with their head, as illustrated in (16):

16) i.

Although these representations are somewhat simplified, they serve to illustrate how the x-complement relation (labelled x) allows the subject to be shared between more than one verb; in (16i), for example “Santa” functions as the subject both of “must” and its x-complement “confess”. As I pointed out in the previous section, WG is unusual in this respect since most other dependency theories impose a maximum of one head per word (Gaifman 1965, Robinson 1970, Mel’čuk 1988). However, the sharing analysis sanctioned by the x-complement relation is valuable in that it yields a unified analysis of so-called ‘raising’ constructions (16ii) and ‘control phenomena’ (16iii). This is particularly useful given that WG recognises neither syntactic movement nor empty categories of any kind, and consequently analyses involving subject NP-raising or PRO are inadmissible.

---

*This is true of Relational Grammar, for example (Perlmutter 1983, Blake 1990). See also Palmer (1994).
One important way in which GRs differ from one another is in terms of their associated semantic content. In WG syntactic and semantic structures are generally isomorphic; in other words there is usually a one-to-one mapping between syntactic and semantic relations. Hudson (1984-a ch. 4, 1990 ch. 7) recognises two core semantic relations, informally described as “er” and “ee”, as well as a range of more specific relations associated with adjuncts such as “time”, “cause” and “place”. The semantic “er” and “ee” relations are usually associated with the syntactic subject and object GRs respectively. Thus in (17) below the subject “Santa” is the ‘steal-er’ of “steal” while the object “the (cash)” is the ‘steal-ee’ (that which is stolen):

17)  

The ‘er’ and ‘ee’ relations are useful as variables in that they define two distinct semantic relations without specifying the precise content of these relations. Thus for example the ‘er’ of “steal” will be an agent, while the “er” of another verb like “faint” will be a patient. Similarly while the ‘ee’ of “steal” is a patient, the ‘ee’ of a verb like “enter” will be more like a locative. Naturally in a passive structure the alignment between syntactic and semantic relations is altered, and the syntactic subject will correspond with the ‘ee’ relation in the semantics. Semantics is a central and complex area of WG, and it is beyond the scope of this thesis to provide a fuller discussion here. Nevertheless, I will return to some of these issues in chapter 4. For now, however, the important point is that each syntactic GR broadly corresponds to a relation in the semantics.

3.3 - WG as a Theory of Knowledge

WG, like PP Theory, is more than a theory of syntax, and might more accurately be described as a theory of linguistic knowledge. In order to comprehend fully WG’s approach to specific syntactic data, such as extraction, it is necessary to understand at least something about the theory’s wider assumptions concerning linguistic knowledge and how it is stored and structured.

86
3.3.1 - Knowledge and propositions

WG takes all human knowledge to be organised as a network of cognitive entities related by a series of propositions (Hudson 1984-a ch. 1). An equivalent conception might thus be that knowledge equates to the set of propositions which serve to link these cognitive entities; we might refer to this set of propositions as \( S_{\text{knowledge}} \) or \( S_K \). Theoretically, according to this conception of knowledge, everything we know could be stored as a proposition contained within the set \( S_K \). Let’s assume, for example, that we know that Prince Philip reads ‘The Sunday Sport’. This fact could then be stored as an individual proposition of \( S_K \) relating our cognitive entry for Prince Philip with our entry for ‘The Sunday Sport’:

18) **Prince Philip reads Sunday Sport.**

The similarity of (18) to a normal sentence of English should not disguise the fact that propositions in WG are formally-defined theoretical units. Like (18) each proposition of \( S_K \) will comprise a connective (predicate) and two arguments. This of course reflects the fact that propositions serve to link two entities within the cognitive network. A more schematic version of (18) is shown in (19) where the two italicised elements represent cognitive entries linked by the connective ‘read’.

19) **Prince Philip** — read —- **Sunday Sport**

Each entity within the cognitive network need only occur once; everything we know about that entity (hence all propositions relating it) will connect to that entity’s single instantiation within the network. So, returning to Prince Philip, within our system of knowledge he has been related to ‘The Sunday Sport’ by the proposition that he reads it. Other things we know about him, such as the fact that he enjoys hunting and that he is married to the Queen of England can also be stored as propositions relating to the same single entity in the cognitive network. In this way hunting and the Queen of England will be indirectly related both to one another and to ‘The Sunday Sport’ via Prince Philip. Figure (20) below shows a simple representation of this pattern; the double arrow between Prince Philip and Queen Elizabeth reflects the fact that they are married to each other:
The 'connectives' in (20) such as "read" and "enjoy" will usually be derived from their own primitive entity in the network. For example, everything we know about reading will be stored under a cognitive entity - 'read'. This entity will then serve as the basis for the connective "read" which links 'Prince Philip' with 'Sunday Sport'. For reasons of simplicity I have omitted further discussion of this here, and I treat "read", "married to" etc. as simple connectives; certainly, there are some propositional connectives, such as "isa" to be discussed later, which cannot readily be derived from any more fundamental cognitive category.

Returning to Figure (20), then, we here see a minuscule fragment of our cognitive network, with propositions corresponding to links between cognitive entities. It isn't difficult to see how this fragment could be expanded into something much larger. For example, everyone else who is known to read 'The Sunday Sport' would also be linked to 'Sunday Sport' by the same connective 'reads', while other things we know about the same paper would also be stored as propositions linking to the same cognitive entry.

3.3.2 - Inheritance and Overriding

Evidently, were our entire knowledge system to be arranged along these lines, with everything we know expressed as a discrete proposition within S_k, then the cognitive network described in WG would be endlessly vast and, to a significant degree, redundant. For example, what if all members of the royal family read 'The Sunday Sport'? This fact would be stored as a separate proposition of S_k, linking the set of members of the royal family with this newspaper. Assuming that we also know that Prince Philip is a member of the royal family, this gives rise to a question concerning the status of our original proposition in (18)/(19) about Philip's own reading habits. Clearly there is an element of redundancy in stating that Philip reads the Sunday Sport, given that we already know that he is a member of the royal family, all of whom read the same thing. Clearly a similar problem will occur whenever a truth is to be stated about an entity which is also true of that instance's more general category.
WG’s solution to this problem lies in a formal system of feature inheritance along the lines of those commonly used in artificial intelligence (Fraser and Hudson 1992). WG assumes that one of the ways in which entities are linked within the cognitive network is by means of specificity. Each entity in the network will be a specific instance of something else. Thus Prince Philip, for example, is a member of the royal family, all of whom are humans, which, in turn, are mammals and so on. This sort of information can be expressed by means of an ontological hierarchy such as (21):

\[
\text{thing} \\
\text{living thing} \quad \text{non-living thing} \\
\text{reptile} \quad \text{mammal} \quad \text{insect} \\
\text{bat} \quad \text{human} \quad \text{Cat} \\
\text{commoners} \quad \text{royals} \\
\text{Princess Margaret} \quad \text{Prince Philip}
\]

This hierarchy is, of course, only partial, and does not necessarily correspond to WG’s classification of things (Hudson 1990 ch. 4-5). Nevertheless (21) does serve to illustrate how each entity in our cognitive network can be arranged on a chain of specificity, all ultimately terminating in the most general category of all - ‘thing’. Since this hierarchy represents part of our knowledge of the world, part of our cognitive network, each connection within it will have to be expressible by a proposition of $S_X$. WG employs a special connective - ‘isa’ - for propositions which link instances with their more general categories. Thus part of the hierarchy in (21) could be expressed by the propositions in (22):

\[
\begin{align*}
i. \text{mammal isa animal.} \\
\text{ii. reptile isa animal} \\
\text{iii. animal isa living thing, etc.}
\end{align*}
\]

Generally what is true of a category will also be true of any more specific instantiation of that category. For example, the fact that all mammals are warm-blooded and viviparous will inevitably hold true of all instances of mammals, such as cats, dogs, humans, and, more indirectly, Princess Margaret. In this way we could say that humans inherit the properties of being warm-blooded and viviparous from the fact that they are mammals. WG formalises this mechanism of natural inheritance and exploits it in eliminating redundancy in $S_X$; quite simply...
anything that is true of Y in (23) will also be true of its instance X:

23) \( X \text{isa} \ Y \)

Any proposition relating to Y will, by default, also relate to X. This allows WG to maximise the effectiveness of any proposition by stating it at exactly the right level of generality; the inheritance mechanism will guarantee that any proposition relating to a category in the ontological hierarchy will also hold true of anything which is an example of that category. Thus returning to the problem raised by the Sunday Sport example, all we need do is state that all members of the royal family read this paper (24):

24) Members of Royal Family read Sunday Sport

By virtue of the hierarchy in (21), Prince Philip will then inherit this property, and no separate proposition is required in relation to him. That does not mean to say that our original proposition in (18)/(19) is invalid or untrue in any sense; instead (18)/(19) is a proposition which need not be stated within \( S_k \), but can instead be derived on the basis of the information in (24). This notion of feature inheritance is extremely simple and, as Hudson (1990 ch. 3) points out, it is implicitly assumed in most theories of language. WG is unusual, though in its explicit adoption and formalisation of the inheritance process.

As we move down the hierarchy in (21), apart from inheriting features from more general categories, each successively more specific entity will also be associated with its own distinguishing properties. Thus the entity ‘human’, being an instance of ‘mammal’, is itself more specific than ‘mammar’, in that it is associated with its own set of properties which serve to distinguish humans from other instances of ‘mammal’, such as ‘cat’ and ‘whale’. In this way, the more specific a thing is, the more individual features it will be associated with. In general, then, entities will combine these more specific features with other properties which are inherited from and shared with more general categories higher up the ontological hierarchy.

In some cases, however, a specific feature of a category will actually contradict a more general property which would otherwise be inherited. For example, what if we know that Prince Andrew reads ‘The People’, rather than ‘The Sunday Sport’? This specific fact about Prince Andrew contradicts what we would otherwise expect of him, given that he is a member of the royal family. In other words, the fact that Prince Andrew reads ‘The People’ rather than ‘The Sunday Sport’ contradicts the generalisation expressed in (24). Nevertheless, it would be a pity to let this single piece of information disrupt a generalisation which would otherwise
hold true of all members of the royal family. What we need to do is to state the specific fact concerning Prince Andrew while simultaneously recognising it as an exception to the valid generalisation expressed in (24). WG handles cases of exception like this by a system of feature overriding (Fraser and Hudson 1992). Quite simply an exceptional property is expressed by two separate propositions, one stating the exceptional property itself and the other specifically negating the more general property which the specific property overrides:

25)  
   i. Prince Andrew reads ‘The People’
   ii. NOT: Prince Andrew reads ‘The Sunday Sport’

The proposition in (25ii) is known as a blocking rule, since its function is to block the inheritance of a contradictory property, thus allowing (25i) to remain valid. The fact that this information is associated with two propositions rather than one helps explain the fact that it encodes exceptional, and thus in some sense marked, information. This system of overriding is a useful way of dealing with exceptions in the cognitive network (S_k), allowing principled generalisations to be maintained, even in the face of contradictory information.

3.3.3 - Linguistic knowledge

According to WG's view, what a speaker knows about his or her language is taken to be a non-discrete part of the network of entities and propositions that corresponds knowledge in general; as Hudson (1984-a p. 6) states:

"... the linguistic network is just part of a larger network, without any clear differences between the linguistic and the non-linguistic beyond the fact that one is about words and the other is not."

In other words, knowledge of language could be said to equate to a set of propositions or rules, call it S_{language}, or S_L, a non-discrete subset of the total set of propositions which make up our cognitive network, S_k. Thus linguistic knowledge will only be distinct from other types of knowledge in so much as it relates to language rather than, say, Prince Philip and his reading habits. This, of course, is a radically non-modular approach to language, and as such it represents a significant departure from much of what is commonly assumed in mainstream linguistic theory (Fodor 1983, Chomsky 1986-b). WG, however, is not alone in rejecting modularity (Lakoff and Thompson 1975, Marlense-Wilson and Tyler 1987, Lakoff 1987, Langaker 1987), and Hudson (1984-a ch. 1, 1990 ch. 4) successfully argues the case for shifting the onus of proof to those who believe that language is formally distinct, separate or
otherwise different from other cognitive systems.

As far as WG is concerned, then, there is nothing inherently different about knowledge of language nor about the way in which it is stored or structured that suggests it should be encapsulated or distinct from any other kind of knowledge. Naturally it is convenient to erect an imaginary boundary around the set of propositions which pertain to language, if only to delimit our area of research. For this reason, then, it is convenient to recognize $S_L$ as a subset of $S_K$. As long as we remember, though, that any boundary between $S_K$ and its subset $S_L$, any boundary between linguistic and non-linguistic knowledge, is drawn purely for the sake of convenience and has no real structural significance.

In accordance with the assumptions outlined above, everything we know about language must be expressed in the form of rules or propositions. The total set of rules relating to a language ($S_L$) will thus constitute the grammar of that language. As we have already seen, propositions are formal units of WG each made up of a connective, or predicate, and two arguments. Hudson (1990) suggests that only three connectives are required for the propositions of $S_L$, ‘isa’ (which I have mentioned already), ‘is’ and ‘has’. Thus all propositions of $S_L$, relating to language will be of the type shown in (26):

$$26) \begin{array}{ll}
i. & X \text{ is } Y \\
ii. & X \text{ isa } Y \\
iii. & X \text{ has } Y 
\end{array}$$

The ‘is’ connective, which should not be confused with ‘isa’, expresses identity between two categories, for this reason it is often substituted by a = sign. For example, the two equivalent propositions in (27) express the fact the subject of “stole” in (17) above is “Santa”:

$$27) \begin{array}{ll}
i. & \text{Subject of "stole" is "Santa"} \\
ii. & \text{Subject of "stole" = "Santa"} 
\end{array}$$

One of the functions of the connective ‘has’ is to express the dependency structure of words and word classes. So the fact that a tensed verb must have a subject can be expressed by the general proposition in (28):

$$\text{It is possible that WG's non-modular views and its rejection of constituency in favour of dependency are linked; while it is relatively easy to find examples of dependency relations in the non-linguistic world (Hudson 1984-a), it is much harder to think of any analogues to endocentric constituent structure.}$$
28) Tensed verb has [1-1] subject

The numbers in square brackets serve to define the minimum and maximum value of the following entity. Thus (28) expresses that a verb can have a minimum of one subject (must have one) and a maximum of one subject (can't have more than one). Where no maximum value obtains, a question mark is used, thus the fact that a word can potentially have any number of adjuncts is expressed in (29), where no minimum or maximum limit is imposed:

29) Word has [0-?] adjunct

This allows a word to have any number of adjuncts, or no adjuncts at all.

The 'isa' connective, as I have already shown, links categories with their more specific instances. Linguistic categories, just like non-linguistic ones, are linked together within an ontological hierarchy. As should be clear from the discussion in section 3.2.1, the principal units of syntax in WG are words and the relations which connect them. Both of these categories are linked on their own hierarchies of specificity:

30) - Hierarchy of words:

31) - Hierarchy of dependencies:

---

This distinction between relational and non-relational concepts, between the word and the relation, is highlighted by Hudson (1990 ch 5) as being of central importance for the description of language.
As with non-linguistic concepts, the hierarchies in (30) and (31) allow WG to express general principles which hold of categories high up in the hierarchy, knowing that these principles, by virtue of feature inheritance, will hold true of all direct and indirect instances of these categories. So, for example, the fact that all words may have any number of adjuncts, as stated in (29) above, will filter down to all categories of words, such as adword, verb and noun, as well as to all instances of these categories, words like "green", "discover" and "ogre". So too the mechanism of feature overriding described in the previous section can also be employed to cancel properties which would otherwise be inherited from more general categories. For example in WG the fact that any verb can have a subject is stored by means of the proposition in (32):

32) verb has [0-1] subject

This allows a verb to have either one subject, in the case of finite verbs, or no subject in the case of non-finites. However, in English the occurrence of a subject with a tensed verb is compulsory, and (32) must be supplemented by two further propositions expressing the specific fact that a tensed or finite verb must have a subject:

33) i. Tensed verb has [1-1] subject
   ii. NOT: Tensed verb has [0-1] subject

(33i) states the specific property of tensed verbs while (33ii) overrides the more general property of all verbs stated in (32).

Note that according to the hierarchy of relations in (31) all words are classified either as pre-dependents or post-dependents. Hudson (1990) argues that this primary classification of words is well-motivated, although he concedes that it is controversial. Certainly it may seem odd to use a word's position in relation to its head - usually itself taken as a reflex of 'deeper' properties - as the basis of any fundamental classification. This is especially true in the case of adjuncts, which may often precede or follow their head with little or no impact on the structure or interpretation of a sentence. Moreover the classification of dependencies in (31) prevents us from defining any single class of adjuncts or arguments; subject and objects, for example, both arguments of the verb, find themselves in opposing sides of the pre-dependent/post-dependent division. I will return to some of the problems inherent with this classification of words in section 3.5.
3.4 - Word Grammar and Extraction Data

Having examined some of the fundamental issues of WG, we are now in a position to explore and evaluate the theory's account of extraction phenomena. Two key aspects of the theory stand out as being of immediate relevance, its monostratalism on the one hand and its rejection of phrase structure in favour of dependency on the other. Just as constituent structure and movement were shown in chapter 2 to be of central importance to PP Theory's account of extraction, so the absence of these two components in WG has equally far-reaching implications for its own account of these data. Just as with every other aspect of grammar, in seeking to account for the extraction data reviewed so far WG will have at its disposal only a single level of syntactic structure composed of words linked by binary dependencies.

3.4.1 - Extraction and long-distance dependencies

Unlike PP Theory, where extraction was described in terms of the relationship between overt and non-overt elements in a syntactic chain, in WG an extracted word will essentially have to bear a relation to a head in the same - or at least similar - way that it would had it not been extracted. Thus to take a concrete example, consider the following examples:

34)  i. Wally will wear a fez.
    ii. What will Wally wear?

In (34ii) the WH-word "what" clearly functions as the object of "wear" and thus must bear an analogous relation to the verb as does "a fez" in (34i). The simplest suggestion would thus be that the two structures are basically parallel with respect to their component relations:

35)

   \[ \text{S} \uparrow \text{X} \downarrow \text{O} \downarrow \text{C} \]
   \[ \text{Wally} \quad \text{will} \quad \text{wear} \quad \text{a} \quad \text{fez} \]

36)

   \[ \text{O} \downarrow \text{X} \uparrow \text{S} \downarrow \text{S} \]
   \[ \text{What} \quad \text{will} \quad \text{Wally} \quad \text{wear?} \]

However, the representation in (36) is inadequate for a number of reasons. For one thing the structure violates projectivity as defined by Hudson (1990, 1994); the dependent "what" is separated from its head "wear" by a number of words which bear no direct relation to either of them. Secondly, the structure in (36) leaves an open dependency between the fronted WH-
word and its head “wear” which remains unresolved throughout the sentence. The size of this open dependency will increase in cases of long distance extraction:

37)

It is generally assumed that open dependencies of this sort lead to processing complexity and tend to be avoided in language (Ninio 1993-b). There is nothing untowardly complex about (36) or (37), however, suggesting that the analysis may be wrong. Another problem with the structure in (36) is that it offers no explanation for the displacement of the WH-dependent, nor, for that matter, for the subsequent inversion of the auxiliary verb and the subject.

Hudson’s solution to these problems is to allow an extracted element to depend on other elements in the clause apart from its ‘normal’ head. This results in the ‘hopping’ analysis shown in (38) where the object relation between “what” and its head “wear” is supplemented by two extra dependencies, represented below the line:

38)

These intermediary dependencies are known as Visitor relations, hence their V-label in (38). Thus as well as being the object of “wear”, “what” is also the visitor of “will” and “wear”. In this way the analysis in (38) relies crucially on Hudson’s assumption, rejected by other dependency grammarians, that a word may have more than one head (see section 3.2.2. and 3.2.4 above for some discussion). To save space, the notation in (39) is often used as a simpler equivalent of (38):

39)
The visitor is described by Hudson (1988-b, 1990) as a semantically empty relation whose function is to mediate between an extractee and its ‘true’ head. The fact that the visitor relation has no semantic content of its own means that it will inevitably coincide with another GR which does have semantic content, such as the object relation in (39); any element which bore only a visitor relation could not be integrated semantically into the rest of the sentence. It doesn’t matter what the other relation is, though; adjuncts, subjects, as well as complements of prepositions can also be visitors:

40. i. V

ii. V

iii. V

The visitor relation takes its name from the nature of its behaviour; in a sense an extractee, in seeking a semantically contentful GR, checks - or ‘visits’ - all potential heads in a structure in the hope that such a relation might be established. The outcome of this process of checking is a visitor relation between the head and the extractee. In this way the visitor relation could be seen as something like a calling card, registering the fact that an extractee ‘passed by’ seeking a (contentful) GR. If, however, no such GR can be established between an extractee and a particular head, then evidently the extractee will have to try elsewhere. In WG this is regulated by a process of transferal; a visitor which has been unsuccessful in finding a contentful relation to a head will be ‘transferred’ to one of the head’s dependents. Thus in (40iii), for example, the extractee “what” first visits the auxiliary verb “did”; since no
contentful relation can be established here, the extractee is transferred to a dependent of “did”, “hide”. Again, no further relation can be established, so once again the extractee is transferred to a dependent of “hide”, the preposition “in”. Here, though, a complement relation can be established between the preposition and “what”; the extractee has thus found its true head, and thus the process of visiting can cease.

From the examples in (38) - (40) we can see then how this process of transferal allows the extra visitor relations to follow the normal dependency structure of the sentences. In each case an extractee is passed down, or transferred, from head to dependent until the point where a semantically contentful relation can be established. This process of transferral, driven by the extractee’s need for semantic integration into the sentence, is the key component in WG’s mechanism for dealing with extraction.

Returning now to our original examples, the presence of the visitor relations in (38)/(39) as against (36) has a number of advantages. Firstly, the structure no longer violates projectivity since the extractee “what” now depends on “will”, the first word that intervenes between it and its ‘true’ head “wear”. Furthermore the fact that the extractee is a dependent of the auxiliary “will” helps explain its displacement to the front of the clause; according to the hierarchy of relations in (31), the visitor is classed as a pre-dependent, and therefore a visitor will precede its head(s). Another benefit of the visitor relations is that they serve to mitigate the effects of the open dependency between an extractee and its ultimate head; in (40iii), for example, the presence of the visitor relations breaks down the long-distance object dependency between “what” and “in” into a series of smaller, more localised dependencies. Here, perhaps, we can begin to see a parallel with PP Theory, in which the presence of intermediate traces in syntactic chains serves to localise long-distance extraction in a similar way. I will say more about this in section 3.5.

3.4.2 - The visitor relation and propositions

Evidently the occurrence and behaviour of the visitor relation, like other GRs, will be determined by rules stated in the form of propositions of the grammar (S₁). One of the most important rules will be the one responsible for linking the extractee with the first finite verb, and thus initiating the process of transferal between heads and dependents, which ultimately allows the extractee to find its true head. Let’s assume, then, that any finite verb may have a visitor. This is expressed as a proposition in (41):
41)  i. Finite verb has [0-?] visitor.
    ii. type of visitor of finite verb = word.

Hudson (1990), borrowing terminology from Gazdar et al. (1985), describes (41i) and (ii) as 'start rules', in that they are responsible for securing an extracted word’s first link in its route to its ultimate head. (41) states that a finite verb may have any number of visitors, of any category. Of course this visitor need not be a WH-word; extracted ‘topics’ too will qualify as visitors of the finite verbs which they precede:

42) The banjo he certainly can’t play _ .

Hudson argues that the rule in (41i), in allowing for a verb to have more than one visitor, accounts for examples such as (43) where the adjunct “last night” is also described as a visitor:

43) Last night what did Wally wear _ _ ?

Returning now to the example in (39), the start rules in (41), in allowing an extractee to be a dependent of the first finite verb “will”, account for the position of “what” at the beginning of the clause since the visitor is classed as a pre-dependent. There is, however, a potential conflict here in that “what” in (39) is also the object of “wear”; objects are classified as post-dependents, and thus “what” should follow its head “wear”. Evidently, then, we must find some way of resolving these contradictory requirements in favour of the visitor relation. In order to guarantee that the visitor relation rather than the object relation will determine the position of the extractee, Hudson (1996) formulates the principle of raising. This states that in cases where a word X has two heads Y and Z, the position of X is determined by the superordinate head; Y will be superordinate to Z if Y is a direct or indirect head of Z. Thus in (39), since “will” is a head of “wear”, the visitor relation between “will” and “what” determines the latter’s position rather than the object relation between “wear” and “what”. This then allows the usual ordering pattern between the head “wear” and its post-dependent “what” to be overridden by means of the blocking rule in (44):

44) NOT: position of post-dependent of head = after it

Once the extractee is linked to the first finite verb, courtesy of the start rules in (41), if no semantically-relevant relation can be established between them, the next task will be to

---

7The italicised ‘it’ in the second part of (44) refers to the italicised word in the first conjunct, here, ‘word’.
ensure that the extractee is transferred to a dependent of the finite verb. The simplest way of achieving this is to allow the visitor of a word X also to be the visitor of X’s dependent.

45) Visitor of word = visitor of dependent of it

Hudson (1990) refers to (45) as a continuation rule in that essentially its function is to allow the extractee to continue searching for a relation. As I mentioned before, the process of transferal expressed by this continuation rule is driven by the extractee’s need for a relation with semantic content, and clearly (45) will apply recursively until a suitable GR is found, allowing the extractee to be integrated semantically into the sentence. The recursive nature of the rule in (45) is illustrated by structures like (46):

46) 

At some point in a grammatical structure an extractee will of course find a proper head, allowing the application of the continuation rule to stop. WG formalises this by means of an end rule; in its simplest form this allows a visitor of a word to bear any other relation to it.

47) visitor of word = dependent of it. (where dependent is any GR apart from visitor)

(47) allows any dependency, adjunct object or subject to be established between a visitor and its head, as witnessed in (40i) - (iii) above. One possibility might be to combine the continuation and end rules in (45) and (47) in one single rule, such as (48):

48) visitor of word = visitor of dependent of it OR dependent of it.

This is a possibility which will be exploited later; for the moment, however, it will be more useful to keep continuation rules and end rules distinct.

Returning once again to the example in (39), between them the rules in (41), (45) and (47) all account for the grammatical extraction of “what”. The start rule in (41) allows a visitor relation to be established between the extractee “what” and the auxiliary “will”. The continuation rule in (45) then applies in allowing “what” to be the visitor of the x-complement of “did”, “wear”. Finally the end rule in (47), is responsible for establishing the object relation between “wear” and its visitor “what”.

100
3.4.3 - WH-word as head of clause

Apart from the visitor relations linking an extractee with various elements in the sentence, Hudson (1988-b, 1990) also analyses WH-extractees as heads of sentences. In this way the tensed verb, which is usually the root of a sentence will be a dependent of a WH-extractee:

49) 

Note, then, that quite apart from the object relation between "what" and "wear", a mutual dependency also exists between "what" and "will", "what" being the visitor of "will", and "will" being the complement of "what". As noted in section 3.2.2, WG is unusual in countenancing mutually-headed constructions such as this whereas other dependency-based theories reject them (Gaifman 1965, Robinson 1970). However, Hudson (1988-b, 1990 ch. 13) presents some good reasons for the analysis in (49) in which the WH-extractee is the head of the sentence. Among these is the fact that certain matrix verbs specify the existence of a WH-clause complement. This would be very difficult to account for if any relation between the matrix verb and the embedded WH-extractee were mediated by the subordinate verb. If, though, the WH-extractee is the head of the embedded clause, then it can be linked as a direct dependent of the matrix verb:

50) 

In addition Hudson (1990) cites the possibility of 'sluicing' where a WH-pronoun occurs without any associated clause:

51) i. I don’t know why he did it.
    ii. I don’t know why.

Once again, this is difficult to explain without a direct relation between the verb "know" and the embedded extractee "why".

101
At first sight it might appear that the additional complement relation between the WH-word and the matrix verb in (49) might suffice to solve the original problem concerning the extractee’s placement at the beginning of the clause; being the head of the finite verb a WH-extractee is predicted to precede this finite verb without needing any ancillary visitor relation. However, the visitor relation in (49) is still required in order to guarantee the transferal of the extractee “what” to its ultimate head “wear”. This is especially true in cases of long-distance extraction such as (46) where the distance between the extractee’s first and ultimate head is even greater. Furthermore, only WH-extractees are analysed as heads of sentences; a topicalised extractee as in (42) will not be the head of the tensed verb, and thus its position at the front of the clause will still have to be determined by the visitor relation alone.

3.4.4 - Constraints on Extraction

The system of start, continuation and end rules described in section 3.4.2 provides a simple but effective way of handling long-distance dependencies and extraction in WG. However, as it stands this system is clearly far too unconstrained, and it is consequently incapable of accounting for the island data and extraction asymmetries described elsewhere. Consider, for example, the subject and complex NP island constraints in (52):

\[
\begin{align*}
\text{i.} & \quad \text{What does chasing tire Mr. Neville?} \\
\text{ii.} & \quad \text{What did you hear rumours that Wally wore?}
\end{align*}
\]

In both of the above examples the continuation and end rules described in section 3.4.2 conspire to allow an extractee to find a proper head within the island domains, incorrectly predicting these structures to be grammatical. In (52i) for example, a violation of the subject island, the extractee “what” can be linked to “does” by the start rules in (41), and then transferred to its dependent “chasing”, thanks to the continuation rule in (45). The end rule in (47) will then apply, allowing “what” to be an object of the gerund “chasing”. The same applies to (52ii), a violation of the Complex NP Constraint; the continuation rule allows the visitor
“what” to be transferred from “did” to “hear” to “rumours” to “that” to “wore”, at which point the end rule can establish an object relation “wore” and “what”.

Evidently what is required is some way of constraining the application of the continuation rule in (45), thus preventing the transmission of the visitor relation in certain circumstances. One way of achieving this would be to state separately that this simple continuation rule cannot apply in certain contexts, such as subject clauses and complex NPs. However, Hudson (1988-b, 1990) suggests an alternative approach whereby restrictions are actually incorporated into the continuation rule itself. In this way rather than allowing a visitor to be transferred to any dependent of a word, we can instead restrict the class of dependents to which a visitor can be transferred. Consider, for example, the revised rule in (53):

53) - Continuation Rule (revised)

Visitor of word = post-dependent of it.

(53) rules out the transferal of a visitor relation to anything other than a post-dependent, predicting that extraction from a pre-dependent will be ungrammatical.

Returning now to (52i) above, as before the extractee “what” starts off as the visitor of “does”. However, the revised continuation rule in (53) ensures that this visitor of “does” can only be transferred to a post-dependent. “Does” has two dependents, an x-complement, “tire”, and a subject, “chasing”. Of these, however, “chasing” is defined as a pre-dependent, even though it follows its head “does”. Recall from the hierarchy of relations in (31) that subjects are classified as pre-dependents, and according to Hudson (1990) this classification is absolute, and will even hold in those cases where a subject actually follows its head (I will return to this point later). In this way, by virtue of the new continuation rule in (53), the extractee visitor of “does” in (52i) will only be transferable to the post-dependent (x-complement) “tire” and not to the pre-dependent (subject) “chasing”. However, no semantically contentful relation can be established between “what” and “tire”, since the latter has its own object “Mr. Neville”. Furthermore, this object is a noun and has no dependent of its own. Thus any further application of the continuation rule in (53) in transferring a visitor relation from “tire” to “Mr. Neville” would be futile. In this way the extractee in (52i) will fail to find any contentful GR and will thus remain dislocated semantically from the rest of the sentence, explaining its ungrammaticality. In other words, no end rule will be applicable in (52i), and thus the object dependency with the bracketed label between “what” and “chasing” will not exist.
Apart from accounting for violations of the subject island, the revised continuation rule in (53) has the added advantage of barring extraction from other pre-dependents, including adjuncts:

54) i. [After meeting Elvis] Noam felt happy
   ii. */**Elvis [after meeting _ ] Noam felt happy

Certainly (54ii) is worse than a structure involving extraction from a corresponding post-adjunct:

55) ℹ️*Elvis Noam felt happy [after meeting _ ]?

Although the revised rule in (53) accounts for the subject island violation in (52i), it is still incapable of ruling out extraction from the Complex NP island illustrated in (52ii). Here the dependency chain between the extractee's first head, “did”, and its ultimate head, “wore”, consists entirely of post-dependents. Evidently a yet more constrained version of the continuation rule is required. Hudson (1988-b, 1990) actually formulates two separate continuation rules in place of the single rule in (53):

56) - Continuation Rules (final version)
   i. Visitor of word = visitor of complement of it.
   ii. Visitor of verb = visitor of post-dependent of it.

According to these rules, only verbs will allow a visitor to be transferred to a post-dependent, as discussed above. Other categories of words, such as nouns, will impose a yet more stringent requirement, allowing a visitor to be transferred only to their complement. Referring back to the hierarchy of relations in (31), one result of this is that extraction from the post-adjunct of a verb will be possible, but impossible from an adjunct of any other word.

Consider once again the example in (52ii); it is now possible to see that the problem with this structure lies in the relation between the noun “rumours” and its adjunct “that”. The transferal of the extractee visitor from “did” to “hear” and on to “rumours” will be unproblematic. According to the continuation rule in (56i), however, “rumours”, being a noun, can only transfer its visitor to a complement, which it does not have in (52ii). Its only dependent here is the adjunct clause headed by “that”. Thus transferal of the visitor relation from “rumours” to its adjunct “that” will be impossible, and the extractee “what” will be stuck with “rumours” as its ultimate head. No semantically contentful relation can be established
between these words, though, and once again the example in (52ii) is ungrammatical because the extractee cannot be linked with the head it requires to be integrated semantically with the rest of the sentence.

The continuation rules in (56) also account for violations of the relative clause island:

57) _____________________________

*Who do you know a man who met

Assuming the structure in (57) to be correct, the relative clause, headed by “who”, is an adjunct of the noun “man” and thus transferal of the visitor from “man” to “who” will be disallowed by the continuation rule in (56i). Quite simply, the extractee “who” will be unable to reach “met” the ‘proper’ head which it requires, and consequently it will be impossible to establish a dependency between these words, such as the object relation marked with a bracketed label in (57).

One of the advantages of the continuation rules in (56) is that, unlike the complex NP constraint in Principles and Parameters Theory, it does not impose a blanket ban on extraction from all dependents of nominals. Extraction from the complement of a nominal is predicted to be fine; in this way we can account for the grammaticality of (58), assuming that the preposition “of” is a complement of “portrait”:

58) Who did he buy a portrait of _ ?

Recall that in PP Theory examples such as these had to be stated as exceptions to the complex NP island (Ross 1967). We can now also account for the contrast in (59):

59) Q: What did the doctor treat the patient with _ ?

A: penicillin/*gangrene.

The answer “gangrene” is inappropriate here since it presupposes that “with”, the head of “what”, is an adjunct of the nominal “patient”, and thus an island for extraction, rather than an adjunct of the verb “treat”.

The continuation rules stated in (56) also allow extraction from the post-adjunct of a verb. This is both useful and problematic for WG, useful in that, unlike PP Theory, WG is able to predict the grammaticality of examples like (60):
60) What did Santa hide the cash in _?

However, in allowing for extraction from adjuncts of verbs generally, these continuation rules fail to account for violations of the adjunct island:

61) i. Sidney became ill [after he ate the sandwich].
   ii. *What did Sidney become ill [after he ate _ ] ?

62) i. Balanescu played ‘Swan Rot’ [because he wanted to annoy his neighbours].
   ii. *Whom did Balanescu play ‘Swan Rot’ [because he wanted to annoy _ ]?

The clauses headed by the complementisers “after” and “because” in (61) and (62) are adjuncts of the verbs “become” and “eat” respectively. In both cases, however, extraction from these adjuncts is forbidden. The adjunct island is actually one of the most fundamental and well-attested of all the island domains, and WG’s failure to account for it is a serious problem for the theory (see section 3.5). Of course one option might be to reformulate the continuation rules in (56) once again, allowing extraction only from complements of all words. Although this would successfully account for the adjunct island, though, there would instead be a problem in explaining the grammaticality of examples like (60) above.

3.4.5 - Constraints on extractability

The end rule stated in (47) allows a visitor of any word to be established as any (other) dependent of that word. This is clearly too unconstrained since it effectively allows any dependent of any word to be extracted, the effects of the revised continuation rules notwithstanding. However, as I have already described in the previous two chapters, constraints do apply on the nature of the extractee itself. For example, just as displacement from the adjunct of a noun was shown to be ungrammatical, extraction of these same adjuncts is also problematic, as illustrated by (63):

63) i. I noticed the marks on the ceiling.
   ii. "*On what did you notice the marks _ ?

A similar effect is also apparent in doctor-type examples:

64) Q: With what did the doctor treat the patient _ ?
   A: Penicillin/*Gangrene.
Once again, “Gangrene” is inappropriate as a response to the question in (64) since it presupposes that the extractee “with what” is an adjunct of the noun “patient”, and thus not extractable, rather than the verb “treat”. To accommodate examples of this sort Hudson (1988-b, 1990) formulates the three end rules in (65) to replace the single rule in (47) above:

65) - End Rules (final version)

i. visitor of word = complement of it.

ii. visitor of verb = post-dependent of it.

iii. visitor of verb = subject of complement of it.

Ignoring for the moment (65iii), the two rules in (65i) and (ii) mirror the continuation rules in (56) in that stricter criteria are imposed on words generally than apply more specifically to verbs; whereas a visitor of a verb can be any post-dependent of it (complement or post-adjunct), a visitor of any other type of word can only end up being its complement. In effect this allows for the extraction of a post-adjunct of a verb, but not a post-adjunct of a noun. Thus we can rule out (63ii) on the grounds that the end rules in (65) will prevent any adjunct relation being established between the extractee “on what” and the noun “marks”. At the same time, however, these rules still allow for the grammatical extraction of an adjunct of a verb, as illustrated in (40i) above.

Turning now to the third end rule in (65iii), consider again (40ii) repeated here as (66):

If this representation is correct then there will be a problem; once the visitor is transferred to “wore”, none of the three end rules in (65) allow for this visitor to be a subject of “wore”; “wore” is a verb, and thus its visitor is predicted by (65ii) to be a post-dependent (complement or post-adjunct) but not a subject. The only provision for subjects in these end rules is stated in (65iii), which allows the visitor of a verb X to be a subject of X’s complement. In this way it will be necessary to modify the representation in (66) as follows:
Here the extractee “who” is transferred only from “did” to “say”, at which point the end rule in (65iii) can apply. Because “say” is a verb, its visitor, “who”, can be established directly as a subject of its object, “wore” (recall from the hierarchy in (31) that the object relation is a more specific instance of the complement relation).

The reason for this seemingly rather complicated account of subject extraction is that it provides an extremely simple and effective account for extraction asymmetries between subjects and other elements. Recall, for example, that the extraction of subjects, unlike objects and adjuncts, is blocked from a clause headed by the complementiser “that” - the so-called ‘that-trace’ filter of Chomsky and Lasnik (1977):

68) i. *Who do you think that _ wore a fez?
   ii. Who do you think _ wore a fez?

Consider now the structures for (68i) and (ii) shown in (69i) and (ii) respectively:

69) i. 

In (69i) the continuation rules allow the extractee to be transferred from “do” to “think” to “that” without difficulty. At this point we should invoke the end rule in (65iii) to establish the visitor “that” as the subject of “wore”, its complement. However, the rule in (65iii) will not apply here since “that” is not an instance of a verb, and thus no subject relation can be secured between its complement “wore” and “who”. Alternatively, of course, once the extractee has been linked to “that” the continuation rule in (56i) could apply instead, transferring the visitor
from "that" to "wore". As I pointed out above, however, there is no end rule allowing for the establishment of a subject relation directly between a verb and an extractee. As far as (69i) is concerned, then, the end rules in (65) prevent any subject relation from being formed between "who" and "wore", accounting for the sentence's ungrammaticality.

Turning now to (69ii), however, when the extractee reaches "think", being an instance of a verb, the end rule in (56iii) can apply, allowing a subject relation to be established between the object of "think", "wore" and its visitor "who". In this way we now have an explanation for the contrast between (69i) and (ii); in the latter, but not the former, the subordinate verb, of which the extractee is supposed to be the subject, is itself a direct dependent of another verb in the matrix clause, to which the end rule in (65iii) can apply without difficulty. The major advantage of this approach is that it effectively subsumes other cases of extraction asymmetries between subjects and objects. In section 1.2.6 of the first chapter I drew attention to the fact that the extraction of a subject from virtually any island gives rise to significantly worse effects than the extraction of an object from a corresponding island domain:

70) i. I know [where Santa hid the cash]
   ii. *What do you know [where Santa hid _ ]?
   iii. **Who do you know [where _ hid the cash]?

71) i. The neighbours banged on the wall [while Balanescu played 'Swan Rot']
   ii. *What did the neighbours bang on the wall [while Balanescu played _ ]?
   iii. **Who did the neighbours bang on the wall [while _ played 'Swan Rot']?

In section 1.3.3 of the first chapter I suggested that rather than looking to any particular property of subjects in order to account for these asymmetries, we might instead seek to extend the 'that-trace' filter to something analogous to a 'complementiser-gap filter' (CGF), where by 'complementiser' I refer to any word which serves to introduce a subordinate clause. In (70iii) and (71iii) the subject 'gap' is adjacent to the complementisers "where" and "while" in the same way as the subject 'gap' in (68i) above is adjacent to "that". Ideally, then, whatever is responsible for ruling out (68i)/(69i) could also be brought to bear in ruling out (70iii) and (71iii) too.

WG is able to exploit this possibility by analysing complementisers like "where" in (70) and "while" in (71) as heads of their respective clauses. In this way the structure of these sentences will be exactly parallel to that of (69i), a violation of the 'that-trace filter'; (72) below offers a simplified representation for (70iii):

109
Assuming that the extractee "who" can be transferred from "do" to "know" and on to its object "where", once again the end rule in (65iii) will fail to apply; "where" is not a verb, and thus (65iii) will be unable to secure a subject relation between its complement "hid" and its visitor "who".

In this way WG is able to utilise the simple end rule in (65iii) to offer a unified account of 'that-trace' effects in (68) as well as the extraction asymmetries illustrated in (70) and (71). This is a significant achievement of the theory, not least because the account does without the complex and convoluted apparatus employed in PP Theory to deal with the same phenomena.

It is unfortunate, however, that WG has no account for the WH-island and adjunct island in (70) and (71) above; although there clearly is a contrast between (70ii)/(71ii) and (70iii)/(71iii), it is nevertheless true that extraction of all elements from these domains is problematic. However, the end rule in (65iii) is only able to explain the ungrammaticality of cases of subject extraction in (70iii) and (71iii). I will say more about this in the following section.

3.5 - Appraisal of the WG Account

To summarise briefly the WG account described so far, the three categories of rule described in the previous sections of this chapter are stated again in (73) - (75) below:

73) - Start Rules
   i. Finite verb has [0-?] visitor.
   ii. Type of visitor of finite verb = word.

74) - Continuation Rules
   i. Visitor of word = visitor of complement of it
   ii. Visitor of verb = visitor of post-dependent of it

---

See, for example, sections 2.4.3 and 2.4.4 of the previous chapter.
In fact Hudson (1990) combines the two continuation rules in (74) and the three end rules in (75) to form just two combined rules shown in (76):

76) - Combined Rules

i. visitor of word = complement of it OR visitor of complement of it.

ii. visitor of verb = post-dependent of it OR visitor of post-dependent of it OR subject of complement of it.

The rules stated in (73) - (76) constitute the basic machinery underpinning WG's account of extraction. The function of these rules is to govern the behaviour of the visitor relation, a structural GR devoid of any semantic content, the presence of which serves to localise the long-distance dependency between an extractee and its true head. The start rules are responsible for attaching the extractee as a visitor of the first available finite verb in the clause, the continuation rules then allow the extractee to be passed down from head to (post-) dependent, establishing a series of visitor relations with other elements in the clause until the end rules can apply, securing a semantically contentful GR between the extractee and a head. This system of rules thus allows a pattern of visitor relations to be built up systematically on the basis of the existing dependency structure, a process driven by the extractee's need for a contentful GR without which it would remain semantically dislocated from the sentence.

This mechanism allows Hudson (1988-b, 1990) to formulate a simple but effective account of extraction based upon a monostratal theory of dependency. In many ways this account is notable for its elegance and parsimony since it successfully accounts for a broad range of data without the complex theoretical apparatus of PP Theory described in the second chapter. Certainly, the WG account outlined above manages without syntactic movement, empty categories, distinct levels of representation like D-Structure and LF, or structurally-defined relations such as c-command, government and L-marking. Indeed, as I noted in the second chapter, these structural relations, as well as the notion of empty category, are in any case closely bound up with the configurational clause structure around which PP Theory is centred. One of the key factors underpinning the relative simplicity of the WG account of extraction is of course Hudson's exploitation of the increased flexibility inherent in a
dependency formalism, especially his use of multi-headed and mutually-headed structures. As I noted in section 3.2.2, there is simply no equivalent of these structures within a constituency-based formalism.

In section 1.4 of the first chapter I suggested that one of the central challenges posed by extraction lay in its apparently contradictory requirements of being unbounded in some circumstances but highly constrained in others. Another key difference between WG and PP Theory concerns how the two models integrate these differing requirements of the data. Recall that in PP Theory extraction was described as an instance of Move-\(\alpha\), a generalised syntactic operation allowing any constituent to move anywhere. Separate principles and constraints such as the ECP and Subjacency were then shown to conspire to restrict the application of Move-\(\alpha\) in such a way that it conformed to attested extraction phenomena. Thus constraints on movement in PP Theory remain entirely separate from Move-\(\alpha\), the mechanism underpinning movement itself. In WG, however, this is not the case; the system underlying extraction is inseparable and indistinguishable from the constraints which restrict its application; very simply, restrictions on what may be displaced and from which domains displacement may occur are incorporated directly into the continuation and end rules, stated in (74) and (75), which govern the behaviour of the visitor relation.

In spite of the various fundamental differences between WG and PP Theory with respect to their treatments of extraction, it is nevertheless interesting to note that the two theories share at least one basic insight; in different ways they both seek to reduce the long distance syntactic relations arising from extraction to a series of more local relationships. In WG this localisation is of course achieved by the multiple application of the continuation rules in (74), allowing a series of visitor relations to link an extractee to a chain of intervening heads. In the case of PP Theory a similar localisation is achieved by means of intermediate traces; instead of moving directly from its D-Structure position to its S-Structure position in one go, an extractee moves in a sequence of shorter ‘hops’ to intermediate positions, leaving a series of traces. The distance between each of these traces is consequently closer, and thus more easy to constrain, than the distance between an extractee and its original source position.

Overall, WG is remarkably successful in offering a relatively simple, monostratal account for a range of extraction phenomena. This is especially notable since WG is actually the only dependency-based theory of syntax to have made any attempt to deal with these data at all. Given, then, this absence of precedent, it is perhaps inevitable that WG will be unable
to account for as broad a range of data as PP Theory, whose own account of extraction has
developed over a far longer period of time and through the work of many more people.

Perhaps the most serious problem facing WG's treatment of extraction is its failure to
account for two of the most basic island constraints, the adjunct and WH-islands. As I pointed
out in section 3.4.4, Hudson's continuation rules do not prevent extraction from adjunct
clauses, illustrated in (77):

77) i. Sid ate a beetle sandwich [because he wanted to impress Polly].
    ii. *Whom did Sid eat a beetle sandwich [because he wanted to impress _]? 

It is true that as it stands the WG account successfully predicts the possibility of extraction
from certain adjuncts, such as prepositions:

78) What did Nigel throw up [in _]?

Nevertheless the effects associated with examples like (77) are sufficiently marked to deserve
an explanation. WG also offers no account for the WH-island:

79) *What do you know why Sid ate 

There is nothing in this structure to prevent the rules in (73)-(76) from applying normally,
predicting that examples such as these should be grammatical. Like the adjunct island, the
effects of the WH-island are clear and well-documented, and WG's failure to account for it is
a serious omission.

In section 2.5 of the previous chapter I noted some problems with the PP Theory
account of the subject island, which attributes the 'islandhood' of subject domains to inherent
structural properties. I argued that this approach fails to predict that extraction from subject
clauses is actually grammatical in a variety of languages when the subject clause itself follows
its head, the matrix verb. Recall too from section 1.3.3 of the first chapter that extraction is
also (more) ungrammatical from an adjunct clause which precedes the matrix verb (see the
examples in (54) and (55) above). It would thus appear that the status of subject clauses as
islands has more to do with their position than any inherent structural properties. Initially the
WG approach seems ideally suited to account for this intuition; the continuation rules in (74) effectively prevent extraction from any pre-dependent, and thus predict (some cases of) the subject island in English, as well as the absence of any extraction in SOV languages like Turkish and Japanese.

Of course, the situation is more complex than this, and extraction is also barred from subject clauses in English WH-interrogatives, which happen to follow their head, the inverted auxiliary verb:

80) *What does [eating _ ] cause heartburn?

As I noted in section 3.4.4, though, according to Hudson (1990), the subject in (80) will still be classed as a pre-dependent, even though it actually follows its head here; the classification of dependents in (31) as pre-dependents or post-dependents is absolute, and is not always necessarily reflected in a dependent’s position in relation to its head. Thus a subject’s status as a pre-dependent in WG actually takes on the characteristic of an inherent feature, something which will hold true irrespective of where it actually occurs. In this sense, then, being a pre-dependent in WG is not so very different from ‘not being L-marked’ in PP Theory. I therefore believe that in spite of initial appearances to the contrary, WG may in fact experience a similar range of problems to the ‘Barriers’ framework in accounting for those cases where extraction is possible from subject clauses which follow their head (see sections 1.3.2 and 2.5).

As I noted in section 3.4.5, one of WG’s strengths is its account of extraction asymmetries between subjects and objects. However WG offers no account for the equally marked and well-attested asymmetries between objects and adjuncts; adjuncts, like subjects routinely give rise to worse effects than objects when extracted from islands:

81) i. I hired a tuxedo [before meeting Sexton Ming at the zoo]
   ii. */Whom did you hire a tuxedo [before meeting _ at the zoo]?
   iii. **Where did you hire a tuxedo [before meeting Sexton Ming _ ]?

The absence of any account for this important asymmetry is a serious omission on the part of the theory.

Another weakness of WG’s account of extraction data, though one it shares with PP Theory, is its inflexibility. In section 2.5 of the previous chapter I drew attention to how extraction of the same word from different islands can give rise to a wide range of effects, from the virtually acceptable to the near uninterpretable (see the five examples in (51) of section
2.5). WG is unable to account for this variation since, in common with other syntactic frameworks, the theory implies an absolute binary distinction between what is grammatical and what is ungrammatical. For example, in section 3.4.4 I described how the continuation rules in (74) explain the effects of the complex NP and relative clause islands. However, no account is offered for the fact that extraction from a complex NP island (82i) is clearly better than extraction from a relative clause island (82ii):

82) i. 'Who did you hear [a rumour that they met _ ]?
   ii. ***Who do you know a man [who met _ ]?

The picture is further complicated by tense; as I described in section 1.3.1 of the first chapter, the presence or absence of tense in an island clause can have a significant effect on its capacity to block extraction. Once again, though, WG has no means of capturing this type of contrast.

This characteristic scalarity associated with island effects is one of the most challenging aspects of extraction. Syntactic frameworks like PP Theory and WG, which, by their very nature, impose fairly rigid criteria of grammaticality and ungrammaticality on language, are simply unable to accommodate the scalar, variable nature of judgements associated with island violations. Indeed, I suspect that no purely structural account of extraction constraints will ever be entirely appropriate for this type of 'ungrammaticality spectrum'. I will investigate this claim further in the fifth chapter, where I will argue that by integrating a theory of dependency syntax with processing-related factors it may be possible to achieve the degree of flexibility required to offer a more realistic account of these data.

There are a number of other miscellaneous problems with the WG treatment of extraction; the theory offers no account of parasitic gaps, for example. Quite apart from data-related issues of this sort, there are perhaps more general questions to be asked about the theory's approach. It could, for example, be argued that in certain respects WG's account of extraction is stipulative and fails to offer any explanation for the data. Why, for example, should it be that a visitor of a verb can be a visitor of its post-dependent but not its pre-dependent? Related to this point is the question of language acquisition; given Hudson's (1984-a, 1990) non-modular, non-innatiast assumptions, virtually everything in language will have to be learnable; a child will somehow have to internalise the set of propositions pertaining to language (S_r) from his or her linguistic environment. It is by no means clear, however, how some of the complex propositions in (73)-(76) could be inferred from observable linguistic data, especially given that language acquisition is assumed to take place only on the basis of
positive evidence. However these questions of acquisition and learnability are complex and controversial issues which lie well beyond the scope of this thesis.

The various problems that WG faces which I have tried to highlight here should not detract from the significance of the theory's account of extraction. What WG proves is that it is possible to formulate a relatively simple, monostratal, dependency-based account of these data without syntactic movement. It is true that WG is unable to match PP Theory in the range of data it covers. However, this is not necessarily due to any limitation of the theory itself, as I have already pointed out, extraction phenomena have received hardly any attention within the context of dependency grammar. Consequently, Hudson's (1988-b, 1990) approach to extraction is, quite literally, without precedent, and is thus unable to build on any existing assumptions. In this respect, then, any direct comparison with PP Theory, which is itself the product of several years' evolution since the Standard Theory, is perhaps a little unfair. However in the rest of this thesis I hope to show how by starting off from the point where WG stops it may be possible to go a good deal further towards the goal of a fully adequate account of extraction.
CHAPTER 4

LICENSING GRAMMAR - TOWARDS A NEW THEORY OF DEPENDENCY

4.1 - Introduction

In the previous chapter I outlined Word Grammar (WG), a dependency-based theory of syntax, and reviewed its account of extraction phenomena. This examination was conducted within the broader context of the second chapter where I presented Principles and Parameters (PP) Theory together with its own approach to the same data. My basic conclusion from the comparison of these two theories was that although WG’s account of extraction is remarkably successful, especially so in relation to other dependency grammars (DGs), it is inadequate in certain key respects, and consequently remains partially inferior to the PP Theory account. As I pointed out at the time, however, WG’s relative inadequacy in this area stems at least in part from the fact that extraction-related issues have not received nearly as much attention in WG and other DGs as they have in the constituency-based linguistic mainstream. The central goal of this thesis is to explore further the potential of a dependency theory to offer a principled and holistic account for the extraction data reviewed in chapter 1, an account which might more realistically aspire to compete on an equal footing with that of PP Theory.

In this chapter I will explore an alternative theory of dependency syntax, which I will refer to as Licensing Grammar (LG). This outline theory will then serve as the foundation for a new dependency-based account of extraction phenomena, to be explored in the fifth chapter. Although LG has much in common with WG, it represents a distinct proposal, and indeed in many ways it marks a significant departure from all other DGs. This is partly due to the fact that dependencies in LG are redefined in terms of licensing relations, and that distinct labelled relations in the form of GRs are eschewed. Section 4.2 introduces the notion of licensing while section 4.3 explores principles of syntactic structure, briefly examining how the licensing relation may also play a part in the morphological structure of words. Section 4.4 then supplements this outline theory of syntactic dependency with a system of derived semantic relations. Although much of this chapter may appear to be far removed from the issues of extraction which are central to this thesis, in fact virtually all the assumptions which I will examine here will turn out to be directly relevant to the account of extraction data to be outlined in the fifth chapter.
4.2 - Dependency as Licensing

4.2.1 - The content of dependency

It is perhaps surprising that one question which, as far as I know, has never been seriously addressed in the Dependency Grammar (DG) literature is what the dependency relation might really amount to. What does it mean to say that one word depends on another? What, in other words, is the precise content of the relation between X and Y in (1)?

1) \[ X \rightarrow Y \]

In the previous chapter I described dependency as an asymmetric binary syntactic relation between a head word and a dependent word. Although most grammarians would agree that dependency embodies an element of asymmetry or inequality, this is clearly too ill-defined to serve as the sole basis of any meaningful syntactic relation; an element of asymmetry or inequality can be discerned between virtually any two words. Outside the realm of language it is usually relatively easy to define the asymmetry inherent in a ‘dependency’ relation, such as that which exists between the United Kingdom and Gibraltar, for example. In this case the asymmetry between the two stems from the colonisation of one entity by another. The dependency in this case can thus be defined economically, politically, and even psychologically. No such easy definitions exist as to the asymmetry which may exist between two words, however, and it is by no means clear in what sense one word is ‘unequal to’ or ‘asymmetric with’ another.

Other than this relatively undefined notion of asymmetry it is difficult to discern any further content to the syntactic dependency relation which is generally accepted among proponents of DG. Tesnière (1959), for example, describes heads as “supérieur” and dependents as “inférieur”, but says nothing as to the nature of words’ respective ‘superiority’ or ‘inferiority’. Robinson (1970), Mel’čuk (1988) and Starosta (1988) also remain silent on this issue. Hudson (1990 ch. 6) does cite certain properties which characterise dependency relations; once again, however, these properties, although sometimes useful as diagnostics to test for the existence of a dependency between words, are largely artefactual, and shed little or no light onto the actual content of the relation itself. I suspect that for most people any content to dependency really resides in the particular grammatical relation (GR) that the dependency instantiates. Indeed, the very presence of these distinct GRs in a theory reflects, to an extent, the absence of content in the more general dependency relation. In this way
dependency often comes to be little more than a blanket term covering a range of diverse relations such as subject, object, adjunct and visitor. This position is most evident in theories such as Relational Grammar (Perlmutter 1983, Blake 1990) and Case Grammar (Fillmore 1968, Anderson 1971, 1977) where individual GRs or case relations have assumed a content and autonomy to the extent that any overriding notion of dependency is rendered largely superfluous. This is one of the reasons for which I referred to these theories in Chapter 3 as first cousins of DG rather than dependency theories in the strict sense (Hudson 1993)\footnote{Rosta (1994) outlines a similar proposal according to which dependency structure amounts to little more than a well-formedness constraint imposed upon a basic system of (more contentful) GRs.}.

In my opinion the absence of a widely-accepted and well-defined content to the syntactic dependency relation has had a detrimental effect on the reputation of WG and other dependency-based theories. A common perception amongst those working within the linguistic mainstream is that dependency is little more than a notational variant of phrase structure. This view, although erroneous (see section 3.2.2 of the previous chapter), is not entirely unjustified, given that grammarians have generally failed to invest the dependency relation with sufficient autonomous content to give it a meaningful character of its own and thus set it apart from other systems of grammatical description. Instead relational linguists often spend much of their time defending the primacy of GRs such as subject and object which are recognised - albeit as derivatives - in constituency-based theories anyway. In this way the debate between relational and non-relational syntax often takes on the characteristics of a separate, though related, debate concerning the primacy or otherwise of a common set of GRs (Blake 1990).

Another unwelcome consequence of the lack of discernable content to the dependency relation is that it leaves DGs susceptible to charges of unconstrainedness. The absence of a well-defined content means that a dependency relation is essentially ‘invisible’, and can consequently be postulated almost anywhere. Say, for example, that we come across a previously unknown phenomenon in a language which seems to suggest that some sort of direct syntactic relation may exist between the subject and the indirect object of a verb. Although unlikely, there is nothing to prevent us from allowing a direct dependency of some type or other between them; there is no inherent property of the relation forcing us to seek an alternative, possibly more principled account of the facts. Ultimately, this can lead to a data-driven approach to language which makes no real predictions and, at its worst excess, hardly qualifies as a theory at all. To recast the issue in more Chomskyan terms, such a theory could
be said to attain observational adequacy too easily at the expense of explanatory adequacy (Chomsky 1965). I will return to the issue of unconstrainedness in section 4.2.3.

4.2.2 - Existential dependency and licensing

What is required, then, in my view, is a syntactic dependency relation with a coherent and meaningful content, over and above ill-defined notions of asymmetry. A relation with this sort of independent content could then serve as the basis of a more dynamic and principled theory of syntax, the relational credentials of which did not reside solely in specific GRs such as subject, object and adjunct. This sort of model might then be more realistically described as a dependency theory, as opposed to a relational theory. Moreover, a better-defined content would make the dependency relation more 'visible', and thus its supposed existence in a given context would be more open to empirical scrutiny. This in turn would make the theory less susceptible to charges of unconstrainedness.

One possibility raised recently by Hudson (personal communication) is that in certain cases a syntactic dependency might amount to some form of contingency relation, whereby the occurrence of a dependent is sanctioned by the presence of the head. Thus with reference to (1) above, the relation between X and Y could express the fact that Y’s existence is contingent on the presence of X. This rather abstract notion of existential contingency can be translated directly into the simpler, more user-friendly concept of licensing. Quite simply we can interpret the relation in (1) as expressing the fact that the head X licenses the dependent Y. In this way it is possible to flesh out the dependency relation into something more meaningful and tangible; one word will, quite literally, be 'dependent' on another for its very existence.

Taking this as my starting point, in this chapter I will explore the possibility that licensing may lie at the heart of all dependency relations, and that consequently each syntactic dependency will have to embody a similar licensing relation between the head and dependent; if one word does not sanction the occurrence of another, then there can be no syntactic relationship between these words. From now on I will use the terms 'licensing relation' and 'dependency relation' interchangeably. So too the terms 'head' and 'licenser' will be used synonymously as will 'dependent' and 'licensee'.

So how, then, might one word license the occurrence of another? Very simply, I assume that a head will license a dependent by formally allowing or sanctioning its presence in a structure. This is stated in (2):
2) - Licensing

X licenses Y in a structure S if X sanctions Y’s occurrence in S.

I will say more about this definition of licensing in section 4.3.6. For the time being, however, I assume that a head will sanction, and thus license, a dependent by actively specifying or requiring its presence; the occurrence of certain words will instantiate the requirement for another, this requirement then serving as a licensing mechanism for the second word. Thus in (1), for example, some lexical property of the head word X specifies the presence of another word, and this specification licenses the occurrence of Y. In this way Y is an existential dependent - or licensee - of X. In many ways a head word’s requirement for a dependent is thus reminiscent of the valency of a predicate, as defined by Chomsky (1981) and Jackendoff (1990) among others, which requires one or more arguments in order to be ‘completed’ or ‘fulfilled’ (Ninio 1993-a). As I will go on to discuss further in section 4.4, though, one of the characteristics of the theory described here is that it maintains a rigid distinction between semantic predicate argument structure and syntactic licensing structure.

When talking about licensing it is important to bear in mind that the existential dependency of a licensee on a head is purely contextual, and only refers to its existence in a particular structure. Perhaps the best way to understand this concept of licensing is by way of analogy; imagine that there is a sport in which a small team of players is selected to play from a larger group before each match. The team coach picks a captain, and the captain then chooses two forwards, these forwards are then each responsible for selecting a defender. In this way a hierarchy of licensing relations, based on selection, can be established between each of the players as illustrated in (3):

3) 

This hierarchy can, of course, also be represented as a WG-type dependency structure:

4) 

In this sort of situation we see how licensing applies purely contextually within the confines
of an individual match; each of the players - whether selected or not - exists independently as a human being as well as a player of the game. However their presence in a specific match is dependent on their being selected, either directly by the coach or by someone else selected by coach. In a similar way each word in a language will exist as an independent entity within the lexicon, although its occurrence in a particular structure is dependent on its being licensed - or required - by another word.

The concept of licensing invests the dependency relation with a specific, relatively well-defined and intuitively plausible content, and one which effectively embodies a requisite degree of asymmetry; a dependent is only present in a structure by virtue of a licensing head. This leaves no ambiguity as to the nature or direction of the asymmetry between them. In fact licensing is a concept which is common to many syntactic theories; Case-assignment in PP Theory, for example, is often described as a licensing procedure by which the occurrence of an overt NP is sanctioned by a head (Chomsky 1986-b). So too Chomsky (1991) states the Principle of Full Interpretation (PFI) as a licensing criterion which holds at the level of Logical Form. Moreover, I suspect that the same concept of licensing is also broadly equivalent to much of what is implicitly assumed in Word Grammar. To take just one example, a proposition stating that word X has a complement is in some sense equivalent to saying that X licenses a dependent. Indeed, it could be that some notion of licensing or existential dependence plays an implicit part in all theories of DG. As far as I am aware, however, no such proposal has ever been made explicit, and until licensing is adopted explicitly and systematically as a component of dependency its presence cannot be formalised, nor can the consequences of its adoption be properly examined and evaluated. Moreover, for reasons which I will describe in section 4.2.3, I believe that much of what is currently taken for granted in most DGs may actually be incompatible with a properly-defined licensing-based conception of dependency.

The remainder of this chapter will be devoted to exploring the possibility of developing a simple theory of dependency syntax around the central concept of licensing. Unsurprisingly, perhaps, I will refer to this outline model of syntax as Licensing Grammar (LG). LG takes up many of the issues raised in the previous chapter, and indeed, the model shares some key assumptions with WG, notably in that both theories are relational, monostratal and lexicalist; LG, just like WG, seeks to express linguistic knowledge in terms of lexical information with

---

2Licensing also plays an important part in many contemporary theories of Phonology (Harris 1994).
reference to a single layer of syntactic structure, without seeking recourse to movement, empty
categories or constituency of any kind above word level (Hudson 1990 ch. 1). The differences
between WG and LG, however, are as important as the similarities, and, as will become
increasingly clear later, many of the suggestions I will make differ in significant ways from the
standard assumptions of WG and other DGs.

4.2.3 - Dependency without GRs

In the previous chapter I described how WG, in common with other dependency-based
theories of syntax, recognises a series of distinct labelled dependencies, or GRs, such as
subject, object, adjunct, x-complement and visitor. This is not a logical necessity however, and
there is no reason why a dependency theory should not do without these relations; in section
4.2.1 I even suggested that to some extent the recognition of distinct GRs as syntactic
primitives is symptomatic of the relative absence of any overriding content in the dependency
relation itself. What I will examine here is the possibility that a more contentful conception of
dependency such as LG, based on the notion of syntactic licensing, may survive, and indeed
even thrive, without having to recognise distinct, labelled subtypes in the form of GRs.

In fact it is true to say that abandoning GRs as syntactic primitives of the theory is not
so much a choice as a necessary outcome of recasting dependency in terms of licensing. To
recognise GRs is to countenance distinct subtypes of the dependency relation. However,
licensing refers to the sanctioning of a word’s existence, and thus to allow distinct subtypes
of this relation would imply that there are different ways in which a word’s existence can be
syntactically sanctioned. This seems unlikely, though; returning to the team sport analogy
illustrated in (3) and (4) above, although I referred to a coach, a captain, forwards and
defenders, the licensing relation between each of them, that of pre-match selection, remains
the same. Similarly, the fundamental biological relation between a parent and child is
unaffected by the sex of the child. In the same way, then, the licensing relation between a head
word and its dependent will remain constant irrespective of the nature or function of the
dependent. In other words, whether Y is a subject, object or adjunct is more a matter for Y
itself rather than the relation between Y and its head X.

There are, of course, clear and important differences between subjects, objects and
adjuncts; this does not mean, though, that these differences necessarily have to be embodied
in distinct syntactic relations. Just as PP Theory was shown to derive differences between
subjects, objects and adjuncts on the basis of phrase structure configurations, so there is no reason in principle why similar distinctions could not also be derived from structural properties of a simple, ‘mono-relational’ licensing-based dependency structure. I will examine this possibility further in sections 4.3.3, 4.3.6 and 4.4. For now the important point to bear in mind is that the added content which licensing brings to dependency is to some extent incompatible with the recognition of distinct labelled dependencies in the form of GRs.

The abolition of GRs as primitive entities in a dependency grammar is a radical suggestion, though one which is not without certain advantages. For one thing, a grammar which utilises only one type of dependency relation will evidently be maximally simple and economical; if we assume that syntax involves relations between words, then the minimal assumption would be that there is only one such relation. Provided it is possible to devise an adequate theory based on a single type of syntactic relation, it will then be up to those who advocate a multi-relational syntax to argue the case for the existence of these different GRs.

More importantly, though, I believe that a viable ‘mono-relational’ theory of syntax such as LG would be more resilient to charges of unconstrainedness than a theory which recognises distinct GRs; I have already described in the previous section how charges of unconstrainedness can arise from the perceived contentlessness of the dependency relation in some theories. However, similar criticisms can also stem from the fact most dependency theories, in recognising GRs as primitive syntactic elements, do not incorporate any inherent limitation on their number or nature; since GRs are not derived from anything else, essentially they constitute an open class which can be expanded as and when required. Which GRs are recognised in a particular theory will be determined more by the data, or even the personal preference of the linguist, than by inherent principles of DG. To take just one example, as I mentioned in section 3.2.4, WG recognises an x-complement relation, where a complement verb shares its subject with its head. There is nothing in the theory, however, to rule out the existence of a ‘y-complement’ relation, which allows, say, for the sharing of an object or adjunct. The fact that such a relation does not exist is essentially because the data do not require it. Once again, then, it might be difficult to formulate coherent predications about language within any such theory when another GR could be just around the corner.

Needless to say, a mono-relational theory of dependency such as LG can sidestep this problem altogether. There is, by definition, only one relation to be recognised, and thus the theory will be automatically constrained. Indeed, in allowing just one syntactic relation LG
could be argued to be more economical and constrained even than the system of Phrase Structure Grammar described in the second chapter. Recall that while GRs such as subject and object are derived in PP Theory, they are derived on the basis of three primary structural configurations: complement, specifier and adjunct (Speas 1990). LG has no equivalent of these three configurations, and recognises only one primary relation, in the form of licensing.

The abolition of GRs in a dependency theory has another welcome consequence in that it entails a reduction in the size of the grammar; as I noted in section 3.2.4 of the previous chapter, every GR that is recognised within a theory is associated with its own distinct properties, each of which, presumably, will have to be stored in the grammar. For example, I described how GRs may differ from one another according to their direction, distribution and morphological marking, and in figure (31) of section 3.3.3 GRs were even shown to occupy their own ontological hierarchy in WG. Quite simply, then, the more GRs that are recognised within a theory, the more facts pertaining to them will have to be stored in the grammar. This is particularly relevant in the case of WG where facts about language are stored as individual propositions. It is not hard to see how the set of propositions corresponding to the grammar, which I described as $S_L$, will have to expand in proportion to the number of GRs.

Conversely, abandoning distinct GRs, along with their associated properties, would result in a reduction in the size of the grammar; since there is only a single syntactic relation, all that need to be stored are properties pertaining to this relation. It could be, of course, that any such reduction is likely to be relatively small in relation to the grammar as a whole. That is not to say, however, that the reduction is insignificant. Quite the contrary, once we eliminate GRs and specific facts relating to them, we are, in essence weeding out virtually all parts of the grammar which can not be stated as lexical information. In fact, as I will argue later, apart from very general facts such as a head parameter setting, a 'mono-relational' syntax can essentially reduce to a lexicon specifying the combinatorial properties of words, a property reminiscent of Categorial Grammars (Ades and Steedman 1982, Wood 1993). In this way LG, by avoiding the need to store information about distinct GRs, is able to match and even surpass the lexicalist intentions of theories like WG (Hudson 1990) and Lexicase (Starosta 1988).

The question of the grammar’s size is not just a matter of theoretical elegance and parsimony; the issue will inevitably have a bearing on questions of learnability and parsing. I will explore some of these issues in chapter 5. Here, though, I hope to have shown how eliminating GRs can have some desirable consequences for LG in allowing the theory to
sidestep certain problems which, to the best of my knowledge, have not been hitherto discussed within the context of DG. The exploration of LG which follows is thus an attempt to offer a coherent and principled account of syntax which makes use of just one structural relation, based on the notion of licensing.

4.3 Licensing Structure

4.3.1 - Some preliminaries

Licensing is a criterion by which the occurrence of a word is sanctioned. Assuming that all words in a well-formed structure are subject to the same sanctioning requirement, we can then infer that all words will need a licenser; any word without a head should simply not appear. This very general requirement can be represented schematically as in (5), where W is a variable over all words, and the downward pointing arrow represents its need of a licenser, and thus its dependence on something else.

\[ \downarrow \]

\[ W \]

The generalisation expressed in (5) is true of all words and thus it does not have to be stated as a property of specific lexical items. Nevertheless, for reasons of clarity I will continue to represent individual words' requirement of a licenser.

Given that licensing constitutes the central dynamic underpinning the system of dependency in LG, we can derive in the simplest possible way the 'no-dangling' requirement of WG discussed in section 3.2.3 of the third chapter; any word which remains unconnected to another word will evidently not be licensed by a head, and thus should not occur. The ungrammaticality of (6) below, for example, repeated from the previous chapter, can be attributed to the fact that "lychee" bears no syntactic relation to anything else and thus is not licensed:

\[ \downarrow \]

\[ \text{Santa stole the cash yesterday lychee.} \]

In this way we can guarantee that any word will have to be linked with at least one other word in the same structure.

---

3The fact that the arrow points from left to right is of no consequence.
Each word's requirement to be licensed will in general be satisfied by other words which are themselves capable of licensing. As I noted in section 4.2.2, one way in which one word can license another is by explicitly requiring its presence; the class of licensing words, which explicitly require, and thus sanction, the occurrence of another word, will include prepositions and transitive verbs as well as complementisers and determiners. Thus, as I noted above, the ability of one word to license another could be said to be equivalent to selecting a complement. Since, however, not all words are licensors, this property cannot be captured by any universal generalisation of the type shown in (5). Instead a word's capacity to license another will have to be stated as an idiosyncratic requirement, analogous to a subcategorisation frame or valency specification. Thus the preposition "with" and the determiner "the" in (7), for example, all licence another word (a property represented by the upward pointing arrow), whereas the noun "gin" does not:

7)  

\[ \text{WITH} \downarrow \text{THE} \uparrow \text{GIN} \]

With the very rudimentary information in (7) we can derive the simple structures in (8):

8)  

\[ \begin{align*}
\text{i.} & \quad \text{the} \downarrow \text{gin} \uparrow \text{with} \downarrow \text{gin} \\
\text{ii.} & \quad \text{with} \downarrow \text{the} \uparrow \text{gin} \\
\text{iii.} & \quad \text{with} \downarrow \text{the} \uparrow \text{gin} 
\end{align*} \]

Note, however, that the same information in (7) will also allow the bad structures in (9):

9)  

\[ \begin{align*}
\text{i.} & \quad \text{*gin} \downarrow \text{the} \uparrow \text{with} \\
\text{ii.} & \quad \text{*the} \downarrow \text{with} \uparrow \text{gin} \\
\text{iii.} & \quad \text{*the} \downarrow \text{with} \uparrow \text{gin} 
\end{align*} \]

Evidently independent constraints will be required to rule out the structures in (9) and others like them. (9i), for example, could be excluded by a simple head parameter stating a general head-before-dependent ordering pattern in English (see section 4.4.7 below). Quite apart from any constraints on head-dependent ordering, (9ii) can be ruled out by an adjacency or locality requirement; the dependency linking "with" with its head will cross the dependency linking "the" with "gin", resulting in a tangling structure. In fact, unlike Hudson (1990, 1994), I assume that all tangling structures are bad and will always result in ungrammaticality. However, I do not believe that this constraint has to be stated as a discrete principle of the

---

4Returning to the parent-child analogy of section 4.2.3, just as everyone must have parents while not necessarily having a child, so each word must have head, or licensor, while not necessarily licensing a dependent.
grammar, and in chapter 5 I will argue that a version of the 'no tangling' constraint can be shown to follow instead from independent processing-related factors.

Finally, the structure in (9iii) could be ruled out by allowing words to impose some sort of categorial restriction on their dependent(s). Thus a determiner such as "the" will only license a nominal like "gin" rather than a preposition such as "with". Quite simply, then, apart from stating the need for a licensee, a head could also specify the category of this licensee. (10) below shows how this sort of categorial information might be represented:

10) \[ \text{THE}^\text{N} \rightarrow \text{x}^\text{N} \]

(10) states that the word "the", itself a nominal, licenses one dependent which is also a nominal. Of course some words may impose more stringent requirements on their dependent, specifying not only the category, but even a specific lexical item. The verb "depend" is a well-known example (Hudson 1995-b); the only possible licensee of this verb is the preposition "on" which, of course, will itself license a nominal:

11) \[ \text{DEPEND}^\text{V} \rightarrow \text{ON}^\text{P} \rightarrow \text{x}^\text{N} \]

If we are to pursue this sort of approach it will be necessary to establish how many categories LG has to recognise. Provisionally we might allow for the four basic categories encountered in most other theories of syntax, N (nominal), P (preposition), V (verb) and A (adjunct) (Chomsky 1970, Jackendoff 1977, Hudson 1990). More complex categories can then be defined on the basis of these basic ones, for example a noun will be classed as a simple nominal (12i) which licenses no dependent, whereas determiners could be defined as nominals which license another nominal (12ii). Similarly a nominal which licenses a finite verb could be classed as a complementiser (12iii).

12) i. \[ \text{BOTTLE}^\text{N} \]
   ii. \[ \text{THE}^\text{N} \rightarrow \text{x}^\text{N} \]
   iii. \[ \text{THAT}^\text{N} \rightarrow \text{x}^\text{V(PN)} \]

However the issue of words' classification and selectional restrictions is not of central importance to the subsequent discussion, and I will not generally continue to represent categorial information of this sort.
Generally, then, words will be licensed by other words which require a dependent. Conversely the capacity of these same words to license dependents will be fulfilled by the presence of other words which, by default, will need a head. In this way a well-formed syntactic representation will consist of one or more chains of dependency with each word being licensed by another. Evidently there must be a point from which these chains originate, since they cannot be of indefinite length. Each structure will thus have to contain one word which, while licensing one or more dependents, will not itself have a syntactic head - the equivalent of the root word in WG. While this would appear to contradict the generalisation expressed in (5), this principle made no claim as to what is actually responsible for licensing a word in any particular circumstance. Although words will generally be licensed syntactically by other words, let’s assume that in any structure one (and only one) word must be licensed non-syntactically, perhaps by pragmatic criteria such as a speaker’s desire to communicate. This ‘root word’ will usually be a matrix finite verb, but there is no reason why any word cannot be subject to this type of non-syntactic licensing:

13)  i. ↓ ii. ↓ iii. ↓ → → →
   Damn!   Hello in the cupboard (in response to a question)

There is nothing to suggest that these structures are not grammatical (in the formal sense), although syntactic theory has seldom paid much attention to them. Unlike WG, then, I am not saying that matrix verbs or examples like those in (13) are headless, but rather I am suggesting that they may indeed be headed - or licensed - though non-syntactically. In this way we can avoid stipulating exceptions to the generalisation stated in (5) that all words must have a head. Following the example of WG, I will continue to indicate the root word in each structure by means of a disconnected arrow.

4.3.2 - Discovering licensing structure

The single word in any structure which is licensed non-syntactically will constitute the ‘root’ of that structure, upon which all other words will depend directly or indirectly. The licensing patterns which result from these chains of dependency will usually, though not always, resemble the dependency structures in WG discussed in the previous chapter; in general a verb will license its arguments and adjuncts while prepositions and determiners will licence nominals, and complementisers will license subordinate verbs.

\[1\] will discuss the licensing of adjuncts in section 4.3.6.
However, one particular advantage of a licensing-based theory of dependency such as LG is that it is usually possible to discover the syntactic structure of a sentence by simple observation, and without having to subscribe to any particular theoretical assumptions; according to the definition of licensing in (2) to find the head of a word Y we need only find the word which sanctions Y’s presence, that is, the word which if removed from the structure would render the presence of Y infelicitous. Here, then, we can begin to discern a degree of ‘testability’ which distinguishes the licensing relation from some other, less contentful versions of dependency. There will, inevitably, be some ambiguous cases, but in general a dependency between two words will be more visible and easier to define than before. Consider, for example, the sentence in (14).

14) Open the bottle quickly.

Turning first to “bottle”, it would appear that the word responsible for sanctioning its appearance is “the”; without “the” the presence of “bottle” in (15i), unlike “it” in (15ii) causes ungrammaticality:

15) i. *Open ... bottle quickly.
   ii. Open it quickly.

Similarly it is evidently the transitive verb “open” which is responsible for licensing the object, in this case “the (bottle)”; with a non-transitive verb such as “walk” the presence of this same object is ungrammatical:

16) i. *walk the bottle quickly.
   ii. walk quickly.

The verb “open” in (14) is also responsible for licensing the adjunct “quickly”:

17) *... The bottle quickly.

Taking all these points together, the only possible licensing structure for (14) will be that shown in (18):

18) \[\text{Open} \quad \text{the} \quad \text{bottle} \quad \text{quickly}\]

In this way we can begin to see how a licensing-based conception of dependency allows us to subject each relation to some form of empirical scrutiny.
While it is relatively easy to pick out the licensing dependencies within a structure, the direction of some of these relations may initially appear to be ambiguous. With reference to the relation between "the" and "bottle" in (14)/(18), for example, while (15) shows that removing the determiner undermines the presence of the nominal, the converse is also true, and the occurrence of "the" without "bottle" is also ungrammatical:

19) *Open the ... quickly.

Could it be, then, that the nominal "bottle" is actually responsible for licensing the determiner? Similarly, while it is true that the transitive verb "open" seems to license the occurrence of the object "the (bottle)", as demonstrated in (16), in some sense this object could also be said to license the presence of the verb; after all, the grammatical occurrence of a transitive verb is to some extent contingent on the existence of a suitable object:

20) *Open .... quickly.

To what extent, then, can we say that "open" licenses "the bottle" when the converse also seems to be true?

There are, of course, well-rehearsed arguments as to why nominals should depend on determiners, and why arguments and adjuncts should depend on verbs (Tesnière 1959, Hudson 1984-a, 1990, Abney 1987, Mel'čuk 1988, Radford 1988); if an object licenses a transitive verb, for example, then what licenses an intransitive verb? However, I believe that it is also possible to resolve these apparent ambiguities simply by appealing to properties of the licensing relation. All that the examples in (19) and (20) show is that dependents need heads and heads need dependents, and that ungrammaticality can result from the absence of one or the other. What we have to distinguish, then, is ungrammaticality arising from the absence of a head, in other words the failure of the 'dependent' to be licensed, and ungrammaticality arising from the absence of a dependent, the non-fulfilment of the head's capacity to license a dependent. In this way it will be possible to determine unambiguously the direction of licensing relations.

In fact I have already briefly referred to two separate properties of the licensing relation which help us identify the head and dependent in a syntactic relation. These two properties are summarised below:

21) i. The head may specify the category of the dependent, but the dependent will never specify the category of the head.

ii. Any dependent, unlike its head, can function autonomously as the root of a structure.
Turning first to (21i), in the previous section I described how a licenser might specify the category of its licensee. For example "the" will require a nominal dependent whereas a complementiser such as "that" or "if" will select a finite verb as a dependent. Conversely, dependents do not generally impose any restriction on the category of their head, and can often be licensed by a head of virtually any category. For example, in (22) below the verb "drink", the preposition "with" and the determiner "the" each specify a nominal dependent:

22)  
   i. drink gin
   ii. with gin
   iii. the gin

However the very fact that this same nominal dependent can be licensed by all three types of head demonstrates that it itself imposes no such categorial constraint on its licenser. The clear implication of this evidence, according to the criterion in (21i), is that determiners, prepositions and verbs are responsible for licensing their companion nominals, rather than *vice versa.*

Turning now to (21ii), in section 4.3.1 above I referred to the possibility of non-syntactic licensing whereby any element can function as the root of a structure. This implies that it should be possible for any syntactic dependent also to occur as a non-syntactic dependent, as the root of an autonomous structure. From this it follows that within each licensing relation the dependent can be identified as the word which, along with its own dependents, can function independently. Returning to the licensing relation between "the" and "bottle", only the latter can realistically function on its own, for example in naming objects (Ninio 1993-a). Determiners such as "a" or "the", however, can never occur independently; for example, even in cases where the use of a determiner is specifically contradicted, its dependent noun must also be included:

23)  
   i. No, (I asked for) the bottle, not a bottle!  
   ii. *No, (I asked for) the, not a!

Once again, then, this implies that determiners - like prepositions - are heads of their companion nouns. Similarly, while a determiner + noun combination may exist autonomously, as demonstrated in (23i), a transitive verb cannot; again, even a verb which is specifically contradicted must be accompanied by its object:

24)  
   i. No, shake it, don't break it!  
   ii. *No, shake, don't break!
Given the object's capacity to function autonomously, coupled with the verb's inability to do so, according to the criterion in (21ii) the conclusion once again must be that the object is a dependent of the verb rather than vice versa.

Here, then, we have two separate criteria which allow us to identify unambiguously the head and the dependent in each syntactic relation. These criteria, in combination with the added visibility that licensing brings to dependency, mean that both the location and direction of each dependency in LG can be discovered more on the basis of empirical observation than theoretical assumption. In this way syntactic structure in LG could be said to be characterised by an unusual degree of transparency.

4.3.3 - Finite and non-finite verbs

So far I have deliberately avoided discussing the licensing properties of finite verbs. Consider now the example in (25):

25) Ted enjoys meths.

What might be the structure for such a sentence? It is tempting perhaps to think that the verb "enjoy" licenses two arguments, here "Ted" and "meths". This would produce a similar structure to the WG-type analyses described in Chapter 3, where direct dependencies link a verb and its arguments. There is certainly a semantic relation between the two arguments and the verb in (25); after all, "enjoy" is a two-place predicate which requires an experiencer and a patient. However, in purely syntactic terms the situation is less clear-cut. There are certainly good reasons to think that the verb licenses its object "meths"; there are, for example, no circumstances in which "enjoy" could occur without an object of some sort. If non-finite, though, the same verb will not license a subject, even though there is still a semantic requirement for one. Thus we find examples such as those in (26), where the semantic properties of "enjoy" are clearly not reflected in the syntactic structure:

26) i. Ted was too drunk to enjoy meths.
   ii. To enjoy meths might be considered odd.

What we can infer from examples such as these is that it seems to be the finiteness of the verb, rather than the verb itself, which is involved in the licensing of subjects, and generally only a finite verb will license all the arguments it requires semantically.
How, then, might we capture the difference between finite and non-finite verbs? Finiteness in some form or other plays a part in most languages (Klein 1994), and the question of how it should be integrated into lexical and syntactic structure has arisen in most theories of grammar. One possibility is to see finiteness as a separate element which, in languages like English, is amalgamated with the verb by a process of grammatical derivation. This is the traditional approach of PP Theory as described by Pollock (1989) and Chomsky (1991), although the lexicalist nature of the more recent Minimalist Programme entails that the combination of verbs and finiteness must be accomplished prior to any grammatical derivation (Chomsky 1993, 1995). An alternative approach is to view finiteness as a basic categorial feature of verbs rather than a separate element. In this way finiteness, or its absence, makes up part of the inherent feature composition of the verb, and has nothing to do with any process of grammatical derivation. This is a view of finiteness embodied in WG (Hudson 1990) as well as Lexicase (Starosta 1988) and HPSG (Pollard and Sag 1987, 1994).

What I would like to explore here is a third view of finiteness, though one which represents a compromise position between the two possibilities outlined above. What I suggest is that finiteness can be analysed as a separate and independent element, though one which is amalgamated with the verb, rather along the lines of the PP Theory analysis. However, this amalgamation or fusion between the verb and the finiteness element has nothing to do with movement or syntactic derivation. Instead, as I will go on to describe in greater detail later in section 4.3.5, certain lexical elements might be described as ‘parasitic’ in the sense that they are required to amalgamate with another, host word, such as a verb, and cannot occur as autonomous words themselves. In this way a finite verb in LG would consist of two amalgamated components, the verb itself and a parasitic finite element, which I shall refer to as FIN. These two components will together constitute a single word, illustrated in (27) below by their inclusion within square brackets:

27) \[\text{FIN V}\]^6

According to this convention a finite verb such as “drinks” would be represented as (28):

28) \[\text{FIN drink}\]^7

---

^6The ordering of the FIN and V elements is irrelevant here, although I will continue to represent FIN preceding the verb.

^7Rosta (1997) has argued for a similar dependency-based analysis of finite verbs.
Of course this sort of analysis raises questions as to the exact nature of FIN, especially concerning how finiteness relates to other categories such as tense, aspect and modality (Comrie 1976, 1985, Palmer 1986). I do not wish to become too involved with these issues here, however, since they represent a significant departure from the central topic of this chapter. Briefly, however, I can see no reason why FIN should not be viewed as a complex element which might comprise information pertaining to tense, aspect, mood and modality, everything, in short, which is not part of the (narrowly-defined) propositional meaning of the sentence (Allwood et al. 1977 ch. 3, Cann 1993 chs. 1 and 8). In this way FIN is partly equivalent to the \(I^0\) node of PP Theory (Chomsky 1986-a) or the M-node of Fillmore’s (1968) Case Grammar. Bearing in mind, then, that FIN might actually function as a convenient amalgam of differing categories, differences in tense, for example, could be registered as distinct indices on the complex FIN element:

29) i. \([\text{FIN}_j \text{ drink}]\) - ‘drinks’
ii. \([\text{FIN}_j \text{ drink}]\) - ‘drank’

The future tense, however, being expressed by an analytic construction in English, would be rendered by the present tense of FIN associated with an auxiliary verb:

30) \(\text{[FIN}_j \text{ will]} \rightarrow \text{ drink}\)

A fuller analysis of the categories which may go to make up FIN lies beyond the scope of this chapter, and for the sake of simplicity I will continue to represent FIN as a unitary category without indexes registering differences in tense or aspect.

Assuming this analysis of finite verbs to be plausible, there is no reason why FIN, as a separate element, should not be free to specify its own syntactic requirements in just the same way as the verb does. Thus both elements of a finite transitive verb could license separate dependents; the verb itself need only license its object, while FIN could be responsible for licensing the subject. Returning now to the original example in (25), we can say that the FIN element rather than the verb “drink” licenses the subject “Ted”.

31) \(\text{[Ted]} \downarrow \text{[FIN enjoy]} \rightarrow \text{ meths}\)

‘Ted enjoys meths’

This analysis has the effect of removing any direct syntactic relation between the verb “enjoy”
and its subject "Ted", a potentially desirable result for a number of reasons. Firstly, there is no longer any need to draw a distinction between the syntactic properties of finite and non-finite verbs; instead this supposed distinction will be a matter of whether or not the verb is fused with a FIN element. Thus in the examples in (32) below the licensing properties of the verb "collapse" remain constant irrespective of whether or not it is fused with FIN:

32) i. Ted [FIN collapse] ‘Ted collapsed’

   ii. Ted [FIN will] collapse ‘Ted will collapse.’

Another advantage of the suggested analysis of finite verbs is that we can now offer a simple account for examples such as those in (26) above where non-finite verbs were shown to be unaccompanied by a subject; given that there is no FIN element associated with these verbs, there will be nothing to license a subject, hence the ungrammaticality of (33):

33) *Ted drink meths
    *‘Ted drink meths’

Conversely, we can also explain why a finite verb must always be accompanied by a subject, even if the verb has no semantic need of one:

34) It [FIN rain] ‘It rained.’

Pleonastic elements such as "it" in (34) appear solely to satisfy the syntactic valency of FIN, and have nothing at all to do with the syntax or semantics of the verb. Consider now (35):

35) Ted [FIN seem] to enjoy a drink ‘Ted seems to enjoy a drink.’

Here once again the subject “Ted” is licensed by the syntactic properties of FIN, although the nature of its semantic relation to the verb "seem" is unclear. There is certainly a semantic relation between “Ted” and "enjoy", however, even though the absence of finiteness associated
with the latter verb means that no subject is licensed. Informally then we could say that the finiteness fused with "seem" licenses a dependent which semantically belongs elsewhere in the sentence. I will say more about structures like this later in section 4.4.4. For the moment, though, we can perhaps see here the beginnings of a possible account of 'raising' phenomena.

Another advantage of the [FIN V] analysis of finite verbs is that it derives a non-configurational equivalent of the syntactic VP; a verb will be more closely bound up with its object than with its subject by virtue of the fact that it directly licenses the former but not the latter. We are therefore in a position to account for the well-known asymmetries between subjects and objects, more specifically the so-called verb-object bonding phenomena (Tomlin 1986, Speas 1990). So too 'surface' distinctions between subjects and objects, such as differences in morphological case marking and word order, can now be derived from the fact that the two are licensed by different elements. In this way we can go some way towards compensating for the absence of distinct subject and object GRs in LG. This whole approach is, of course, not dissimilar to that of Principles and Parameters theory, where the I® head of the derived [I + V] constituent is responsible for Case-assigning (hence licensing) the subject while the V® head assigns Case to the object. The differences between these analyses, however, are as important as the similarities. For one thing FIN, unlike I®, does not correspond to a structurally-defined position and is consequently entirely absent in non-finite constructions. Another crucial difference, of course, as I have already pointed out, is that the [FIN V] word is neither the product of movement nor is itself dependent on movement for 'feature-checking'. Instead [FIN V] represents the lexicalised fusion of two elements which enter into a word-internal relation. The nature of this word-internal relation will be the subject of section 4.3.5.

4.3.4 - LG and Categorial Grammar

What I have suggested so far is a simple system of dependency syntax which involves just one structural relation. This relation remains constant and invariant, irrespective of the two words which participate in it. Thus a determiner or a preposition will license a dependent in exactly the same way as does a verb or FIN. Similarly all words will have the same need of a head, no matter what that head actually is in a given structure. As I mentioned briefly at the end of section 4.2.3, one of the benefits of this 'mono-relational' system of dependency without distinct GRs is that it allows virtually the entire syntax to be stored in terms of lexical entries; since there is only one syntactic relation in LG, all that needs to be stated is how
individual words participate in this relation. Essentially, then, except for very general properties of the licensing relation, such as its direction, virtually all of the LG syntax will thus reduce to a lexicon listing words and their licensing properties; individual syntactic structures will merely be a reflex of words' propensity for licensing others. In this way the syntactic structure in (36) is a direct and transparent product of the lexicon fragment in (37):

36) 
\[ \text{We [FIN know] that Ted [FIN crash] his Bentley} \]

\[ \text{\textquoteright We know that Ted crashed his Bentley\textquoteright} \]

37) 
\[ \text{KNOW, CRASH, THAT, FIN, HIS, TED, WE, BENTLEY}^8 \]

The licensing information in (37) is entirely non-directional, and thus the ordering of elements in (36) will have to be achieved by separate linearisation principles, possibly in the form of a head parameter (see section 4.4.7). Apart from this, though, there is no additional information in (36) which is not contained - or could be contained - in (37); I have already discussed how categorial constraints on dependents can be incorporated into lexical entries (see examples (10), (11) and (12) in section 4.3.1). Similarly, while some form of adjacency or locality constraint is clearly operational in (36), as I mentioned previously, I believe that this can be derived from independent processing factors, and consequently need not be stated as part of the grammar. I will discuss this further in chapter 5.

The reduction of virtually all of the syntax to the lexicon, as indicated in (37), is very much in line with the lexicalist intentions of WG (Hudson 1990 ch. 1) and Lexicase (Starosta 1988). Indeed, as noted in section 4.2.3, by abandoning distinct GRs a theory such as LG is actually able to store a greater proportion of grammar in terms of purely lexical information than other theories are able to achieve. In many ways the ultra-lexicalist orientation of LG is also reminiscent of Categorial Grammar (CG) (Ades and Steedman 1982, Buszkowski et al. 1988, Oehrle et al. 1988, Wood 1993), particularly the earlier and relatively more straightforward combinatorial systems of Bar-Hillel (1953) and Ajdukiewicz (1935); the licensing properties expressed by the information in (37) are, to all intents and purposes, equivalent to combinatorial requirements; to say that word X licenses a dependent is equivalent

\[ ^8 \text{Of course the fragment in (37) could also serve as the basis for other structures such as "we know Ted".} \]
to saying that X must combine with a dependent to form a well-formed structure. In this way the mutual satisfaction of a head’s need for a licensee and a dependent’s need for a licensor (38i) can be viewed as a version of the Application rule in CG, illustrated in (38ii):

\[
\begin{align*}
\text{(38i)} & \quad \text{X} + \text{Y} \implies \text{X} \text{ Y} \\
\text{(38ii)} & \quad \text{X}[\text{Y} + \text{Y} \implies \text{X}}
\end{align*}
\]

In both (38i) and (38ii) the entities X (or X[Y) and Y are brought together on the basis of their combinatorial requirements, the outcome being a well-formed structure in which these requirements are satisfied.

However, I should point out that important differences remain between the ‘monoreminal’ theory of dependency instantiated in LG and Categorial Grammar. For one thing, while the satisfaction of a head’s requirement of a dependent in LG is reminiscent of Application in CG, most versions of CG include several other rules such as Composition, Raising and Associativity (Lambek 1958, Cohen 1967, Wood 1993), for which LG has no equivalent. Moreover, it is also important to bear in mind that Application in CG is a purely combinatorial operation, and has nothing to do with existential dependency. Consequently the inherent element of asymmetry which characterises licensing relations in LG is absent in the rules of CG. Another important difference between Categorial Grammar and LG is that in the latter, like most other dependency theories, a word’s dependency structure is assumed to remain distinct from its category. Thus both “enjoy” and “die” will be classed as verbs in spite of their different valency specifications. In CG, however, the differing combinatorial requirements of these two verbs is reflected in two separate categories; whereas “enjoy” would be of the type S[N,N], “die” would be classed as a S[N (Lambek 1958, Ades and Steedman 1982, Hudson 1990 p. 109).

4.3.5 - Words and elements

Above I suggested that finite verbs could be analysed as single words consisting of two separate components, the verb itself and a finiteness element FIN. Although I suggested that dependency theories such as WG have no need of any constituents larger than the word, constituency of some sort will obviously have to be recognised at a sub-lexical level. Words

---

9The parallels between Dependency Grammar and Categorial Grammar are reflected in the theory of Compiler Grammar (Ninio 1993-a), a procedural model of syntax which incorporates insights both from WG and combinatory logic.
are undoubtedly composed of smaller units such as morphemes and phonemes, and all theories must have a means of expressing this fact. The potentially problematic aspect of the analysis of finite verbs in section 4.3.3, though, is that both elements of the word were shown to specify their own distinct syntactic properties; both FIN and the verb are capable of licensing their own dependent. This analysis would appear to fly in the face of the earlier claim that words were the only units of syntactic relevance for a dependency theory such as LG. How, then, can sub-lexical elements such as FIN enter into a syntactic relation which is supposed to be the preserve of words?

Another important question raised by the same analysis concerns the nature of the relation between the two elements of a finite verb; is the relation between FIN and the verb morphological, and, if so, how does this morphological relation differ from the syntactic licensing relation examined so far?

The simplest answer to both of these questions would be to suggest that there is actually no difference between the syntactic relations which link words and the ‘morphological’ relations which link elements within words; the same licensing-based version of dependency could serve to link both words and elements. In this way the two elements comprising a finite verb would enter into a licensing relationship with one another while at the same time licensing their own dependent words:

\[39) \text{Wally} \ [\text{FIN hate}] \rightarrow \text{syntax} \]

\[\text{‘Wally hates syntax’} \]

According to (39), FIN is responsible for licensing the verb rather than vice versa, and thus FIN will constitute the root element of a matrix finite structure\(^{10}\). There are many good reasons supporting this analysis. For example, if instead the verb were the head of FIN, then we would be forced to draw a distinction between the licensing properties of ‘finite’ verbs, which licensed FIN, and ‘non-finite’ verbs, which did not, thus cancelling one of the key benefits of introducing the FIN element in the first place. Moreover, there are certain words, such as complementisers like “that” and “if” which specifically select a finite verb as a dependent. The only plausible way to formulate such a constraint is to take finiteness itself, in the guise of FIN,

\(^{10}\)This analysis with FIN as the head of the verb is broadly concurrent with other theories which assume a distinction between finiteness and the verb form. In PP Theory, for example, VP is a complement of P head, the traditional location of finiteness features (Chomsky 1986-a).
to be the direct dependent of the complementiser, and then allow the verb to depend on FIN.

Although WG does not share this analysis of finite verbs, in certain circumstances it does allow sub-lexical components to participate in dependency relations, notably in the case of clitic constructions (Hudson 1984-a, Volino 1990) and gerundives (Hudson 1990). The latter are said to be composed of two elements, a verb and an ING 'clitic'. These two elements enter into a dependency relation, yielding a word with the internal structure shown in (40):

40) [walk ING] - 'walking'

In essence, then, all I am suggesting is that this kind of analysis could be extended to finite verbs as well.

Hudson (1990) takes the rather unusual view that both the bracketed constituent in (40) and the two elements which constitute it should be classed as words. In this way a single word such as "walking" can be made up of two (or more) smaller 'words', "walk" and "ING". This analysis raises some awkward questions for WG, and might represent the thin end of a large and dangerous wedge; in what sense is a word the largest unit of syntax if the same word can be a part of a larger word? Why, moreover, can't any two words, such as "hate" and "syntax" in (39), themselves constitute another 'word'? I will avoid problematic questions such as these by examining an alternative approach to the issue whereby elements, rather than words, are recognised as the basic units of syntax.

Let's assume, then, that in LG the basic unit of syntax is the element which, for our purposes, could be considered broadly equivalent to the morpheme. Licensing dependencies will, therefore, hold exclusively between elements:

41) E1 E2 E3

Words, on the other hand, are lexical rather than syntactic units; they represent those elements (or combinations of elements) which could be said to have acquired phonological autonomy. Words are thus derivative components which are only of relevance to syntax indirectly, via the elements which go to make them up. More often than not there will, of course, be a one-to-one correspondence between elements and words, in cases where a word is composed of a single element. This will be true, for example, of nouns, prepositions, determiners and 'bare', non-finite verbs. In these cases licensing relations could be said to exist between words, but again,
only indirectly, by virtue of the fact that these words happen to be elements. Thus the w-
labelled brackets in (42) indicate those elements which exist as words:

(42) i. \[E1]_w [E2]_w [E3]_w ii. [in]_w [the]_w [box]_w

In other cases, however, a word could be composed of two or more elements which enter into a licensing relation (43i). I assume, however, that two elements which do not enter into a licensing relation will never amalgamate, and thus will not be able to constitute a word (43ii):

(43) i. \[E1 \ E2]_w ii. *\[E1 \ E2]_w

As far as English is concerned, this type of analysis seen in (43i) need not apply only to finite verbs and gerundives, but might also extend to genitive nouns as well:

(44) \[GEN \ Ted]_w [Bentley]_w

'Ted’s Bentley’

In these cases, then, a word could be said to correspond to the fusion of two elements which enter into a syntactic relation. While some elements, such as “Ted” and “drink” can appear as autonomous words, other elements, like FIN, can only ever occur when fused with another element. In this sense these elements could be described as ‘parasitic’, in that they require a host in order to be lexicalised. The important point is, though, that the licensing relation will remain the same regardless of whether or not the two participating elements are fused together as a single word. Thus the three structures in (45) below are all structurally equivalent, differing only in the combinations of elements which have been lexicalised.

(45) i. \[E1]_w [E2]_w [E3]_w ii. \[E1 \ E2]_w [E3]_w iii. \[E1 \ E2 \ E3]_w

This raises a number of interesting possibilities. For one thing, it is possible that licensing structure actually remains fairly constant universally, with languages differing according to which elements, or combinations of elements, are stored as words. Thus, for example, while FIN and the verb are fused in English the same two elements may occur as two separate words in another language:
This appears to be true of languages like Lao (Hoshino and Marcus 1981) and Fijian (Schütz 1985), for example, where finiteness and tense are routinely expressed as distinct particles entirely separate from the verb. Similarly, while (44) represents a genitive noun in English consisting of a nominal and a parasitic genitive element GEN, in a language like French the genitive element occurs as a separate word, as, indeed, it may in English too, in constructions such as "the eye of the storm".

The other possibility, of course, is that separate single-element words which enter into a licensing relation in English may occur as one fused word in another language. Indeed, the prediction is that amalgamation might potentially occur between any head and its dependent. Without wishing to go into too much detail here, in certain languages it is possible to find potential 'word-internal' analogues of virtually all the syntactic dependencies examined so far. For example, it could be argued that 'pro-drop' languages with agreement inflection on the verb allow the three-way fusion of FIN with a verb and a subject pronoun. Similarly, languages like Albanian and Swedish with definite marking on nouns might be displaying fusion between a parasitic definite determiner element and its nominal dependent, while the rich system of nominal inflection in Finnish and Hungarian could be the result of fusion between what in other languages are adpositions and their nominal dependents. In the same way object clitic constructions in languages like Modern Greek could be analysed as the fusion of a verb and an object pronoun, while another possibility is that languages might allow the amalgamation of a verb and its full nominal dependent. This could be what lies behind some of the incorporation constructions discussed in Baker (1988). One thing that is predicted never to occur, of course, is fusion between two elements which do not enter into a licensing relation.

Evidently much work remains to be done in exploring some of these ideas more fully. I can only hope to have offered an insight here into how the licensing relation in LG allows for the possibility of breaking down the sometimes rather artificial distinction between syntax and morphology (Spencer 1991 chs. 6 - 9). If what I have suggested above is correct, then any distinction between the two will essentially be a matter of where a language superimposes lexical structure on the system of primitive morpho-syntactic licensing relations between elements. Much of the above discussion is really only an extension and formalisation of what is currently assumed in WG (Hudson 1990, Rosta 1997). So too the supposed continuity
between morphology and syntax in LG echoes some phrase structure theories, where configurational structure is also taken to extend beyond syntax to the internal morphological structure of words (Williams 1981, Selkirk 1982).

4.3.6 - Arguments and adjuncts

Examples of the licensing relation discussed so far have involved the mutual participation of the head and the dependent; a dependency has been seen to arise from the simultaneous requirement of a head to license something and the (universal) requirement of a dependent to be licensed. This is true, for example, of all the syntactic relations in (36) above, as illustrated by the accompanying lexicon fragment in (37). We might refer to this type of relation as mutually-sponsored or 'active' licensing in that the licensing relation between the head and dependent is a product of the syntactic requirements of both.

Although all the licensing relations discussed so far have been active in the above sense, it is natural to ask whether all instances of the licensing relation have to be mutually-sponsored in this way. Evidently it is unlikely that a dependency could exist if neither the head nor the dependent required it, since any such relation would be totally unmotivated and untestable, and I will discount this possibility here. Nevertheless, there is no logical reason why a dependency could not arise from the lexical requirements of just one of the two participating elements. For example a licensing relation could be sponsored solely by the head. This would involve cases where a head X licenses a dependent Y which happens to be licensed already by another head Z, and consequently doesn't need to be licensed by X. This would result in the sort of double-headed structure illustrated in (47) below:

47)

Following Hudson (1990), I assume that structures corresponding to (47) may indeed occur in natural language, as in 'headless relatives', for example. However I will have nothing more to say about this here. Instead I would like to concentrate briefly on the other possible permutation of the licensing relation, where a dependency is specified solely by the dependent and not the head. I will refer to this as dependent-sponsored licensing, or 'passive' licensing.

Cases of passive licensing will arise when a syntactic relation is the product only of the dependent's need for a licenser. In other words the head will not actually require a licensee, but will merely allow one. The definition of licensing in (2) states that a head X licenses a
dependent Y by sanctioning Y’s presence. As I have already described, one way in which X can do this is by actively requiring Y’s presence. However, there is no reason why X should not also sanction Y’s presence passively without actually requiring it. This sort of situation is illustrated in (48i) below where, according to the lexical information in (48ii), the head X does not specify any syntactic requirement to license a dependent. From now on I will identify this sort of dependent-sponsored licensing relation by means of an arrow with a dotted stem:

48)  

As an analogy to this situation we could speak of a whale ‘licensing’ the presence of barnacles upon its body without necessarily needing - or even wanting - them there. The relationship between whales and barnacles is driven solely by the latter’s need of somewhere to sit. Essentially, then, the distinction between active and passive licensing boils down to whether a head actually requires a licensee or merely allows one.

The relevance of active (mutually-sponsored) and passive (dependent-sponsored) licensing is that together they allow us to derive in a simple and principled way a distinction between arguments and adjuncts (Dowty 1982, McGonnell-Ginet 1982). If we had to isolate a single defining characteristic of arguments it would surely be that their presence is required by their head (Matthews 1981 ch. 6, Huddleston 1984 ch. 5). If, then, a head specifically licenses its argument and that same argument requires a licensor, it is easy to see that a head-argument relation will inevitably be mutually-sponsored, and thus active in the sense described above. The presence of adjuncts, on the other hand, is not specified by a head. For the most part they can occur freely and in any combination, their presence basically determined by the speaker’s wish to include them (Huddleston ibid). All that is required of an adjunct is that it has a head - a property stated of all words in the generalisation in (5) above. Since, then, the relation between an adjunct and its head will be driven solely by the former’s need to be licensed, the mere presence of another word, indeed any word, in a structure should suffice to license an adjunct passively.

What I am thus suggesting is that in LG arguments can be defined as active licensees and adjuncts as passive licensees. However it is very important to stress at this point that the definition of arguments and adjuncts in these terms absolutely does not amount to the recognition of distinct syntactic GRs. As I have already suggested, GRs constitute an extensionally-defined, and potentially open set of relations which can be added to on a fairly
ad hoc basis. Passive and active licensing, on the other hand, simply represent two of the three logically possible permutations of the binary licensing relation; a relation between X and Y can be initiated by X, or Y or both X and Y. In appealing to these differing permutations to derive a distinction between arguments and adjuncts, it is possible to compensate once again for the absence of distinct GRs in LG, while avoiding the various pitfalls associated with their recognition, described in section 4.2.3.

Consider now the representation in (49) where, once again, the passive licensing relation between the head “wreck” and the adjunct “yesterday” is distinguished by a dotted arrow stem:

49)

Ted [FIN wreck] his Bentley yesterday

'Ted wrecked his Bentley yesterday'.

A structure such as this will be the product of the lexicon fragment in (50):

50)

FIN WRECK HIS TED BENTLEY YESTERDAY

FIN, “wreck” and “his” all specifically require licensees. Whatever fulfils this licensing capacity of these words in a given structure will be classed as their syntactic argument. Thus in (49) “Bentley” is the argument of “his” which in turn is the argument of “wreck” which is itself an argument of FIN. All these relations are mutually-sponsored. However in (49) “wreck” is shown to license two dependents while only specifying one licensee in (50). Evidently, then, one of these relations in (49) will have nothing to do with the head, and will have to be sponsored solely by the dependent. This dependent will therefore be classed as an adjunct.

An important implication of this definition of arguments and adjuncts in LG is that it entails a contextual rather than an absolute categorial distinction between the two. Essentially an element Y will be classed as an argument if there happens to be a head X to actively license it, and as an adjunct if there is no such head. In (50), for example, there is no inherent syntactic property of “yesterday” which defines it as an adjunct rather than an argument; it requires a head in just the same way as any other word. “Yesterday” happens to be an adjunct in (49) simply because there is no active valency specification available to license it. More precisely,

---

11Following the discussion in the previous section, I assume that FIN licenses two dependents, one of which will be the verb with which it must be fused.
by the time “yesterday” occurs all active licensing requirements have been satisfied. In other circumstances, however, “yesterday” could be defined as an argument:

51)  

\[ \text{the party [FIN be] yesterday} \]

‘The party was yesterday’.

Assuming that “be” licenses its complement, the licensing relation between it and “yesterday” will be mutually-sponsored in (51), and thus active. In this way “yesterday” can be classed as an argument or an adjunct according to its licensing context. Similarly, a phrase headed by a preposition can also serve either as a syntactic argument or adjunct, depending on the licensing environment:

52)  

i.  

\[ \text{Ted [FIN see] the bottle on the table} \]

‘Ted saw the bottle on the table’.

ii.  

\[ \text{Ted [FIN put] the bottle on the table} \]

‘Ted put the bottle on the table’.

The only difference between (52i) and (ii) arises from the different selectional properties of the two matrix verbs; “see” licenses one argument while “put” licenses two. Here then we can see a possible advantage of deriving ‘adjuncthood’ contextually.

However, this approach does raise certain important questions. Why, for example, can’t “yesterday” be the argument of any head:

53)  

*Ted wrecked his yesterday?

Similarly, referring back to the ungrammatical example in (6) in section 4.3.1, why can’t the word “lychee” be passively licensed like an adjunct rather than being left to ‘dangle’?

54)  

*Ted [FIN wreck] his Bentley yesterday lychee.

*I ‘Ted wrecked his Bentley yesterday lychee’.

\[ ^{12} \text{I will examine this issue further from a processing point of view in the fifth chapter.} \]
Syntactically, I suspect that there may be nothing wrong with (53) or (54) at all. Certainly in both examples the licensing properties of all the constituent words are satisfied. Instead the problem seems to be a matter of interpretation (see section 4.4.6 below). For example, “lychee” in (54), unlike “yesterday” in (49), bears no meaningful relation to the rest of the sentence. Similarly the incompatibility between “his” and “yesterday” in (53) would appear to be rather more to do with meaning than syntax and licensing. In this way the problems associated with examples like (53) and (54) can be attributed to semantics rather than syntax, and thus need not detract from the contextual definition of arguments and adjuncts that I am proposing here.

This issue of semantic structure in LG is to be the topic of the following section. The important point for now, though, is that a distinction between arguments and adjuncts can be derived from two of the three logically possible permutations of the licensing relation. One advantage of this approach is that it is no longer necessary to recognise a primitive adjunct GR. Moreover, the contextual definition of arguments and adjuncts that I am suggesting here is of particular importance to LG since, as I will show in the next chapter, it may help account for some of the extraction asymmetries described in the previous chapters.

4.4 - Syntactic and Semantic Relations

I have suggested that instances of LG’s syntactic licensing relation arise from something akin to words’ combinatorial properties, their individual requirements to license and be licensed by other words. In the previous section I described how syntactic arguments could be defined as elements whose licensing relation with their head was sponsored by combinatorial properties of both the head and the dependent, while adjuncts could be defined as elements whose licensing relation with their head was a product of the valency properties of the dependent alone. However this definition of arguments is purely syntactic, and does not necessarily have anything to do with semantic predicate-argument structure. In this section I will examine the relationship between syntax and semantics in LG, and explore the possibility of deriving a rudimentary system of semantic relations on the basis of the syntactic licensing structures described so far.
4.4.1 - The mismatch between syntax and semantics

Consider the licensing structure in (55) below:

(55)

\[
\text{Ted} \quad \text{[FIN want]} \quad \text{to} \quad \text{drink} \quad \text{something}
\]

'Ted wants to drink something'.

Although the syntactic structure here is well-motivated, what this representation fails to illustrate is the fact that "Ted" is interpreted as the subject both of "want" and "drink". These verbs are two-place predicates and, semantically at least, they require two arguments, one of which is "Ted". In spite of this, however, there is no syntactic relation linking "Ted" with "want" or "drink" in (55); as I have already discussed, there is good reason to think that the finiteness associated with "want" is alone responsible for licensing "Ted", and that the two verbs license just one dependent each, their respective objects "to (drink something)" and "something". For this reason the WG analysis discussed in section 3.2.4, which allows two verbs to 'share' the same syntactic subject, is inapplicable here.

Examples such as (55) point to an important discrepancy between syntactic and semantic argument structure within LG; while "Ted" is a semantic argument of both "want" and "drink", it is a syntactic argument - an actively-licensed dependent - only of FIN. What is evidently required, then, is some means of supplementing the representation in (55) with a semantic structure which would, for example, express the fact that "Ted" is an argument of both of the verbs. Following the assumptions of WG, discussed briefly in the previous chapter, I assume that semantic structure can be expressed by a parallel system of relations, as illustrated in (56) below:

(56)

\[
\text{Ted} \quad \text{[FIN want]} \quad \text{to} \quad \text{drink} \quad \text{something}
\]

The lower set of arrows point from semantic predicates to semantic arguments, just as the upper set point from syntactic heads to syntactic arguments. However, the superficial parallels between the two structures here should not be allowed to mask the fundamental differences between them. For one thing the lower arrows do not represent dependency relations since they have nothing to do with licensing. Instead the semantic relations in (56) could be interpreted as pointing out the arguments of a predicate. More helpfully, perhaps, each of these
lower arrows could be taken to represent the assignment of a semantic role by the predicate to an argument. In this way “drink”, for example, will assign one role to “Ted” and another to “something”. This system of role assignment, similar to θ-role assignment in PP Theory (Jackendoff 1972, Chomsky 1986-b), can then serve to express predicate argument structure in the semantics. Describing the lower set of arrows in (56) as roles is especially useful since it helps avoid confusion by maintaining a distinction between syntactic structure, described in terms of licensing relations, and semantic structure, described in terms of roles.

While the dual system of relations/roles illustrated in (56) may be reminiscent of the parallel dependency structures of WG (see example (17) of section 3.2.4), it is important to bear in mind that unlike the WG example, the syntactic and semantic structures here are not isomorphic. One thing that is clear, however, is that whatever the mismatch between syntax and semantics embodied in (56), there must be a principled and uniform relation between the two. Were this not the case, there would be no sure way of deriving a sentence’s meaning from its syntactic form. At the very least, then, a sentence’s semantic structure will have to be derived from its syntactic structure according to some sort of systematic and principled algorithm. This will be the topic of the remainder of this chapter.

Before going on to examine some of these issues in greater detail, it is necessary first to clarify briefly some points of terminology; I have described “Ted” in (55) as the subject of “want” and “drink”, and “to (drink something)” and “something” as the objects of these verbs, this in spite of the fact that in section 4.2.3 I specifically argued against the existence of GRs such as subject and object. In fact, as should be clear from the above discussion, I am using the terms ‘subject’ and ‘object’ here informally to refer to semantic roles rather than syntactic relations. In this way the roles in (56) above could be labelled as follows:

Of course it might be wrong to use the terms ‘subject’ and ‘object’ in this somewhat unconventional way, referring to semantic roles, especially given the plethora of different meanings already attributed to them. It would actually be possible to use any two general labels

---

13This distinction between syntactic relations and semantic roles corresponds to a similar distinction made by Palmer (1994).
for the roles assigned by a predicate to its arguments, such as R1 and R2, or even ‘er’ and ‘ee’,
the semantic relations of WG discussed briefly in the third chapter (Hudson 1984-a ch. 4, 1990
ch. 7). However, it seems to me that this might actually lead to further confusion; after all,
everyone recognises “Ted” as the ‘subject’ of “want” and “drink” in (57), and “Ted” clearly
bears a semantic role in relation to both of these verbs. All I have done, then, is apply an easily-
recognisable and fairly general label to this semantic role, making no assumptions as yet as to
its precise content.

In the rest of this section I will sketch a very general outline account of semantics
centred around just two basic or ‘primary’ semantic roles, which I will continue to label
informally as ‘Subject’ and ‘Object’. I assume that these two primary roles, like the ‘er’ and
‘ee’ relations of WG, have virtually no inherent content, and that their only real function is to
remain distinct from one another. Thus just as the term ‘positive’ often has no real content
other than as the opposite of ‘negative’, so I believe that ‘Subject’ may have no inherent
content other than as a counterpoise to ‘Object’. The contentlessness of these two roles then
allows their precise nature to be determined on an individual basis by the assigning predicate.
Thus, for example, while the Subject of a verb such as “faint” will be an experiencer, the
Subject and Object of “drink” will be an agent and patient respectively:

58)  

Ted [FIN drink] something

<agent> <patient>

In this way the two primary semantic roles in (57) and (58) can be thought of as empty vessels
which are filled by the (precise) role assigned by a predicate. One of the advantages of this
system is that there is no need to recognise a large and potentially unconstrained class of
precise semantic roles as primitive entities of the grammar, instead these roles will be the by­
products of the interaction of predicates with the two primary roles Subject and Object:

59)  

i. FAINT + Subject = <experiencer>

ii. DRINK + Subject = <agent>

iii. DRINK + Object = <patient>

The precise mechanisms underpinning this mapping procedure are not immediately relevant to

14Capitalised so as to help avoid confusion with subject and object GRs in other theories.
the current discussion, however, and the remainder of this chapter will concentrate instead on how the recipients of the two primary semantic roles are identified on the basis of syntactic licensing structure.

4.4.2 - Identifying Objects

From the examples discussed so far it should be clear that if a predicate has more than one semantic role to assign then one of these - what I refer to informally as the Object - will be assigned to a syntactic dependent of the predicate itself. In (56), (57) and (58), for example, the objects of “want” and “drink” are their respective licensees “to (drink something)” and “something”. I assume then that the Object of a word can be routinely identified as its syntactic dependent. This principle is stated in (60):

60) - The principle of Object identification

Object of X is a syntactic licensee of X.

In this way the object roles, once again represented beneath the structure, can be derived on the basis of (55) as follows:

61)

Although all Objects are defined as dependents in (60), it is clearly not the case that all syntactic dependents will qualify as Objects; according to the assumptions outlined above, Subject and Object refer to the two primary roles assigned to different semantic arguments of a predicate. Although “to” in (61) has a syntactic dependent, “drink”, it seems unlikely that this dependency has any reflex in the semantic predicate-argument structure of the sentence. Certainly it is hard to conceive of “to” assigning a role to “drink”, and the latter will thus not be an Object of the former. For similar reasons “want” will not qualify as an Object of FIN.

Nevertheless, the fact that a syntactic relation has no corresponding semantic role does not necessarily mean that it is irrelevant to semantic structure. For example, “to” in (61) is the licensee of “want”, and therefore, according to the principle in (60), it alone will be assigned the Object role by “want”; however from an interpretative point of view it seems more likely

---

\[^{15}\text{I will generally use the terms 'Subject' and 'Object' to refer to the primary semantic roles as well as to the recipients of these roles.}\]
that the real argument of "want" is the entire phrase "to drink something". In this way we might allow the Object role assigned to "to" to permeate down through its syntactic dependents, to "drink" and "something". This could be achieved by a process of percolation:

62) - The principle of percolation

if X assigns a semantic role R to Y, then R will percolate to all direct and indirect syntactic dependents of Y.

The effect of (62) is to derive a system of semantic phrasing, whereby a role assigned to an element will also be borne by all syntactic dependents of that element. In this way few, if any, syntactic dependencies will ever be entirely irrelevant to semantic structure.

4.4.3 - Subjects and d-command

As far as Objects are concerned, then, the principle stated in (60) guarantees that there will be an absolute correspondence between semantic and syntactic structure, the effects of the percolation principle in (62) notwithstanding. The recipient of the Subject role evidently cannot be identified in the same straightforward way, though, given that verbs do not even enter into a direct licensing relation with their Subject arguments. The Subject will thus have to be located on the basis of some rather more complex aspect of syntactic structure. In fact, I believe that Subjects can be uniformly identified by reference to a relatively simple structural relation which I will refer to as d-command, a dependency-based analogue of c-command in PP Theory:

63) - D-command

An element X will d-command its sister Y, along with any direct or indirect dependents of Y.

The term 'sister' in (63), borrowed from phrase structure terminology, refers to any two elements which share the same head. Consider now the schematic licensing structure in (64).

64)

In (64) A and C qualify as sisters since they share the same head, B. A and C will thus d-command each other. According to (63), A will also d-command all direct and indirect dependents of its sister C - D, E, F and G. Of course C will not d-command D, E, F or G since these are all its own dependents. D and E, which share the same head, C, are also sisters and
will consequently d-command each other. In addition, D will also d-command E's direct and indirect dependents, F and G. D will not d-command A or B, however, since these are neither its sisters nor dependents of its sisters. In other words, just as with c-command in phrase structure theory, the condition of 'sisterhood' in (63) prevents the d-command relation from extending 'upwards' beyond an element's head. Note, though, that any element can be d-commanded by more than one other element; for instance E, F and G in (64) are all d-commanded both by D and A.

The similarities between d-command and c-command, discussed in section 2.3.2 of the second chapter, should be obvious; the only real difference between the two is that d-command is formulated in terms of dependency rather than constituent structure. However, unlike c-command in PP theory, I believe that d-command can be utilised to identify the recipients of Subject roles in a principled and uniform way on the basis of a syntactic licensing structure. I assume that a predicate's Subject can be identified as the nearest d-commanding argument:

65) - The principle of Subject identification

Subject of X is the nearest d-commanding argument to X

The reference to 'nearest' in (65) - which I will refer to as the 'proximity condition' - can be measured in terms of dependency links. Put simply, the fewer dependency relations separating two elements, the nearer they will be to one another. Thus F in (64), for example, is d-commanded by two elements, A and D. Of these, however, D is closer to F than A is, since the two are only separated by three dependency relations, between D and C, C and E and E and F. A and F, however, are separated by four relations, between A and B, B and C, C and E and E and F. For this reason, then, D rather than A will qualify as a Subject for F. If, on the other hand, C had a Subject role to assign, the only possible recipient would be A, since this is the only element which d-commands C. Thus in seeking the Subject of an element X we simply move 'up' the dependency structure through X's direct and indirect heads. At the first point in which one of these heads has a second argument (not an indirect head of X) then this second dependent will qualify as the nearest d-commander, and hence Subject, of X.

Consider now the simple syntactic structure in (66):

16It is true that verbs in PP Theory are c-commanded by their subjects. However this is a consequence of the fact that subjects are defined as specifiers and these c-command heads; c-command as such has no formal part to play in the identification of subjects in PP Theory.
The predicate "want" assigns a Subject and Object role; according to (60) the Object role will be assigned to its licensee "a". The principle of percolation in (62) ensures that the Object role assigned to "a" will also be transferred to its dependent "drink". According to the principle of Subject identification in (65), in order to find the Subject of "want" we simply locate the closest d-commanding argument. In this case there is only one possible candidate, "Ted"; FIN, the head of "want", has a second dependent, "Ted", which will qualify as the nearest, and indeed only, d-commanding element to "want". "Ted" will therefore be assigned a Subject role by "want":

Now that the primary semantic roles assigned by "want" have been located, it is up to specific lexical properties of the predicate to determine the precise content of these roles.

The system of Subject identification and d-command described above is, as far as I know, entirely new within the context of dependency grammar, and, as such, represents a distinct characteristic of LG. Nevertheless, the principle in (65) does bear certain superficial similarities to the definition of subjects in other theories. For example in Predication Theory (Williams 1980, 1983, Napoli 1989) the subject of X is defined as "the single argument of X that is located outside the maximal projection of X" (Williams 1983 p. 287). Implicit (curiously so perhaps) in this definition is the understanding that the subject will be the closest argument to X outside X's projection. Similarly, the 'proximity condition' in LG's principle of Subject identification may initially appear to be reminiscent of the (somewhat archaic) Minimal Distance Principle of the Standard Theory (Rosenbaum 1967). However, this principle never sought to derive or identify Subjects as such, but instead determined the referential interpretation of the empty category PRO. Consequently the criterion of minimal distance actually held not between a predicate and its Subject, as I am suggesting, but instead between a non-overt subject pronoun and an overt nominal argument situated in a different clause. In contrast, however, as I will go on to discuss below, LG's principle of Subject identification can...
serve as a reliable axiom for locating semantic Subjects over a wide range of syntactic structures, from simple matrix clauses to more complex constructions involving 'raising', 'subject and object control' and 'ECM' structures.

4.4.4 - 'Raising' and 'subject control' structures

The inclusion of the 'proximity condition' in the principle of Subject identification in (65) allows a degree of flexibility to be built into the location of Subject arguments. This is a positive advantage in that it potentially allows for a uniform axiom of Subject-identification to be maintained over a variety of otherwise diverse syntactic structures. Moreover, unlike an object, which is located in absolute terms, a single word can simultaneously qualify as the 'nearest d-commanding argument' - and hence the Subject - of more than one predicate at a time. Consider once again the sentence in (55) repeated below:

55')

Ted [FIN want] to drink something
'Ted wants to drink something'.

We can now employ the principle in (65) to identify the Subject of both "want" and "drink". The Subject of "want" will obviously be located in exactly the same way as it was in (67). In the case of the verb "drink", however, in searching for the closest d-commanding argument, we look first to its head, "to". This has no other dependent which could d-command "drink" so we must move one step up the dependency chain to the head of "to", "want", which also has no second dependent. However the head of "want", FIN, does have a second argument, in the form of "Ted", which will thus qualify as the nearest d-commander to "drink". In this way "Ted" is identified as the recipient of the Subject role of both "want" and "drink":

68)

Ted [FIN want] to drink something

Here, then, we can see how the principle of Subject identification in (65) can apply in a principled and uniform way to a simple finite clause, as in (67), and a 'subject control' structure like (68). In fact exactly the same principle of Subject identification also applies to 'raising structures' involving verbs like "seem":

156
I am unsure whether “seem” really assigns a Subject role to “Ted”, or if the latter is only a syntactic dependent of the FIN element fused with “seem”. Essentially, though, this is unimportant here; the verb “like” clearly does assign a Subject role to “Ted”, and thus we should expect “Ted” to be identified as the Subject of “like” according to the principle in (65); the first head up the licensing chain from “like” which actively licenses a second dependent is, once again, FIN. This same dependent, “Ted”, will therefore qualify as the nearest d-commander, and thus Subject, of “like”.

Apart from the question concerning the semantic relation between “Ted” and “seem”, the structure in (69) is exactly parallel to that in (68), in spite of the fact that these two constructions often receive very different analyses in other theories. In PP Theory, for example, the structure in (69) would be said to involve the movement or ‘raising’ of “Ted” from its D-Structure position in the subordinate clause to its surface position as the subject of the matrix clause. (68), on the other hand, would not be the product of movement, but would instead require a non-overt PRO element in the subject position of the subordinate clause, whose interpretation is ‘controlled’ by “Ted” in the matrix clause. Moreover, the analyses in (68) and (69) require no additional syntactic dependency relations, such as the x-complement and extra subject GRs which would be required in a WG analysis of these constructions (see the examples in (16) in section 3.2.4 of the previous chapter).

4.4.5 - ‘Object control’ and ‘ECM’ structures

The principle in (65) may also be used to identify the recipients of Subject roles in other constructions. (70) below is an example of an ‘object control’ structure:

70)  
  Ted [FIN persuade] Fred to drink the petrol

‘Ted persuaded Fred to drink the petrol’.

The verb “persuade” is semantically a three-place predicate which licenses two syntactic dependents, one nominal and the other clausal. I assume, then, that “persuade” will assign an
Object role to each of these, as well as a Subject role to “Ted”, identified according to the principle in (65). A partial semantic structure for (70) is shown in (71):

71) Ted [FIN persuade] Fred to drink the petrol

Turning now to the predicate “drink”, as before its object will correspond to its dependent “the (petrol)”, in accordance with (60) and the percolation principle in (62). Moving up the licensing chain from “drink” to locate its Subject, we pass from “to” to “persuade”. Unlike the previous examples, the verb “persuade” licenses a second dependent, “Fred”, and this will therefore qualify as the closest d-commander and Subject of “drink”. In this way one of the Objects of “persuade” will also receive the Subject role of “drink”, a case of ‘object control’:

72) Ted [FIN persuade] Fred to drink the petrol

Although this semantic structure may seem complex, from an interpretational point of view it appears to be correct, and, moreover, it is entirely derived from the much simpler syntactic structure by the two principles in (60) and (65).

I assume that exactly the same analysis will apply in similar cases of ‘object control’ with verbs such as “help” and “ask”. Essentially, the proximity condition of the Subject identification principle allows the difference between ‘subject control’ structures such as (68) and ‘object control’ structures like (72) to be reduced to a matter of whether or not the matrix verb licenses a second syntactic dependent apart from the complement clause. If it does then this second dependent will automatically qualify as the nearest d-commander to the non-finite verb in the embedded clause, and will thus receive its Subject role. If there is no second dependent, then the nearest suitable argument to the embedded verb will automatically be the licensee of the matrix FIN, as in the case of (68) and (69). There is, however, one notable and troublesome exception to this generalisation involving the verb “promise”.

---

17 The case of “promise” is also cited by Jackendoff (1972) as an argument against Rosenbaum’s (1967) Minimal Distance Principle.
is interpreted as the subject of the complement verb "buy", this in spite of the fact that "promise" has a second licensee, "Ted" which qualifies as the closest d-commanding argument to "buy" and thus, according to (65), should receive its Subject role:

73)

'I promised Ted to buy some gin'.

I cannot pretend that I have a satisfactory answer to this problem. One possibility is that "Ted" is neither a syntactic argument nor semantic Object of "promise" in (73), and could instead be analysed as an adjunct; as I will describe later in section 4.4.6 this would entail a very different syntactic structure from that illustrated in (73), accounting, perhaps, for its specific interpretation. Whatever the case may be, though, it is worth pointing out that "promise", as far as I know, is a unique exception.

There is one final construction worth examining in connection with the principle of Subject identification, illustrated in (74):

74)

'I expect Ted to drink it'.

This is an example of what is sometimes referred to as an ECM (Exceptional Case Marking) construction (Chomsky 1981). For our purposes, though, (74) differs from the 'object control' structure in (72) only in terms of the semantic argument structure of the matrix verb. I assume that like "persuade", "expect" here licenses two syntactic dependents. Unlike "persuade", however, "expect" semantically is only a two-place predicate; the only thing 'expected' in (74) is the state of affairs whereby Ted will drink something. I assume, then, that "expect" assigns a Subject role and only one Object role, to the clause headed by "to". The other syntactic licensee of "expect", "Ted", will thus not be assigned any semantic role by the verb, another apparent mismatch between syntactic and semantic argument structure:

75)
Clearly "Ted" will have to bear some role in (74) and (75), otherwise it will remain semantically dislocated from the rest of the sentence. It so happens, of course, that "drink" assigns both a Subject and an Object role. The latter will be assigned to its dependent "it" and the Subject will once again be defined as the nearest d-commanding argument. Looking up the licensing chain from "drink", the first head with a second argument is the verb "expect" whose hitherto roleless dependent "Ted" will be identified as the Subject of "drink":

I hope then to have demonstrated how the syntactic licensing structures of LG described in section 4.3 may serve as a basis for deriving semantic structure in a systematic, principled and relatively simple way. In particular I have shown how the structural relation of d-command along with the principle in (65) can correctly identify the recipient of a predicate's Subject role in a variety of circumstances; a predicate’s Subject will be identified in exactly the same way for finite verbs (67), ‘subject control’ and ‘object control’ complement structures (68) and (72) as well as ‘raising’ and ‘ECM’ constructions (69) and (76)\(^\star\). In all these cases a predicate’s Subject role is assigned to its nearest d-commanding argument, as stated in (65). Moreover, the simple analysis which underpins all these examples has not involved any syntactic movement or the recognition of empty categories such as PRO. Neither has the analysis required any specific syntactic relation such as WG’s x-complement. The sharing analysis sanctioned by the x-complement relation in WG is now an automatic reflex of the flexible principle of Subject identification. At the beginning of section 4.4 I referred to the inherent disparity between syntax and semantics in LG, a disparity first introduced by allowing FIN rather than verbs to license subjects. What I hope to have shown, then, is that this disparity, in allowing a relatively simple and uniform semantic structure to be created on the basis of a diverse set of syntactic constructions, may actually be a strength of the theory.

\(^\star\)In fact, as I will go on to show later, the same definition of the subject relation also allows us to account for ‘control’ phenomena in non-finite adjunct clauses.
4.4.6 - The syntax and semantics of adjuncts

In section 4.3.6 I suggested that adjuncts could be defined contextually as elements which are ‘passively’ licensed by their head. Thus in (77) below the adverb “yesterday” is classed as an adjunct since “drink” licenses it passively, without specifying any requirement to do so. In other words “yesterday” is the sole sponsor of its relation with its head.

Ted [FIN drink] aftershave yesterday

‘Ted drank aftershave yesterday’

As I pointed out before, however, in defining adjuncts in this way LG faces a potential problem, since there is nothing in principle to rule out an ungrammatical structure such as (78):

*Ted [FIN drink] aftershave lychee

*Ted drank aftershave lychee’

What prevents “lychee” from being licensed passively here, and thus being an adjunct of the verb “drink” in the same way as “yesterday” in (77)? As I briefly mentioned at the end of section 4.3.6, I suspect that syntactically - that is according to principles of licensing - (78) may actually be fine. Instead the problem with this example appears to be one of interpretation. More specifically, “lychee” bears no meaningful relation to the rest of the sentence in (78). The key to ruling out this sentence would thus appear to lie in the semantic roles rather than the syntactic relations which link adjuncts with their heads, and in this section I will explore briefly what these semantic roles might be.

In seeking to establish a semantic link between an adjunct and its syntactic head, one possibility would be to recognise a series of semantic roles along the lines of those encountered in WG and many other theories. Thus as well as being syntactic adjuncts, “yesterday” in (79i) and “in Bucharest” in (79ii) below could also bear some sort of ‘Time’ and ‘Place’ role respectively:

79) i. Ted [FIN drink] aftershave yesterday

TIME
Similar semantic roles could then link predicates with a whole range of adverbs, such as cause, manner and purpose.

However, there are a number of problems with the analyses shown in (79). For one thing, even theories which do not otherwise recognise syntactic GRs will have to acknowledge roles such as 'Time' and 'Place' as primitives at some level, even if, as I suggest, they might be semantic rather than syntactic relations. This might be uncomfortable for a mono-relational theory such as LG, however, since, as Anderson (1977 ch. 1) notes, it could be construed as admitting distinct GRs - albeit disguised as semantic roles - through the back door. Of course it might be possible to derive the precise content of roles such as Time and Place from the adjunct itself. Here, though, there is an untidy contradiction; in the case of Subject and Object it is the semantic predicate - the assigner of the role - which is responsible for determining its precise content, whereas in the case of the putative 'adjunct roles' it would instead have to be the recipient of the role, the adjunct, which defined their content.

There is a yet more serious problem with the representations in (79); these analyses do not really make sense within the role-based system of semantics that I am exploring here. As I have already mentioned in section 4.3.6, adjuncts, by their very definition, are peripheral elements whose presence in a sentence is determined largely by the speaker's wish to include them. However a predicate's capacity to assign a semantic role is an individual, lexically-specified property, and is not achieved on an ad hoc basis. How, then, could "drink" and "buy" coherently assign Time and Place roles to adverbs whose very presence is, to all intents and purposes, random? If "drink" were to assign a Time role to "yesterday" in (79i), for example, then the implication would be that "drink" always assigned a Time role; this would be extremely uneconomical given that in the vast majority of cases "drink" will occur without any temporal adverb. More to the point, even if "drink" were somehow to assign a role to "yesterday", then why should it not also assign a similar role to "lychee"? Above I suggested

\[\text{Note here that the preposition "in" is taken to assign a semantic Object role to its syntactic dependent "Bucharest". This assumption, or some version of it, is common to most theories.}\]
that the ungrammaticality of examples such as (78) resided in the semantic properties of words like “lychee” as opposed to words like “yesterday”. However, it is by no means clear how the semantic properties of these words could be distinguished, given that they are both taken to be the recipients of roles assigned by other words; it would certainly be very difficult - if not impossible - to formulate a plausible constraint preventing certain words from being assigned certain roles. Instead it would seem far more natural to formulate constraints on role-assignment in terms of the assigner rather than the recipient.

There is a relatively straightforward, though perhaps somewhat unorthodox, solution to these problems; quite simply, my suggestion is that instead of being assigned a semantic role by the verb which corresponds to their syntactic head, adverbs may instead themselves assign a semantic role to this same verb. In other words, I believe that these adjuncts may constitute semantic predicates, which take their syntactic head, or whatever they serve to ‘modify’, as their argument. This is illustrated in (80):

80)  

This analysis solves all of the problems raised above; since “yesterday” now assigns a role to “drink” rather than vice versa, we can allow this role to remain inherently contentless, just like the Subject and Object roles discussed above. Its precise temporal content in (80) can then be determined by properties of the predicate “yesterday” in exactly the same way as the true nature of the Subject role assigned to “Ted”, for example, is determined by the predicate “drink”. The implication of this, of course, is that just as with ‘agent’ and ‘patient’, we no longer have to recognise specific roles such as ‘time’ or ‘place’, since these will simply be by-products of the interaction between individual predicate-adjuncts, such as “yesterday”, and a (contentless) semantic role. Similarly, we no longer have to stipulate that verbs assign an extra semantic role in cases where adjuncts happen to appear; just as “drink” consistently assigns two semantic roles, Subject and Object, so “yesterday” will also consistently assign its own role, thus guaranteeing its own semantic link with a sentence. In this way it is also possible to

20This view of adjuncts as predicates partly corresponds with that of Categorial Grammar (CG), according to which adverbs are analysed as functors of the type (S|N)(S|N), which take a verb (S|N) as an argument (Ajdukiewicz 1935, Wood 1993). However, the LG analysis here, unlike CG, maintains that adjuncts are syntactic dependents of verbs rather than vice versa.
rule out ungrammatical examples like (78); it is simple to state as a lexical property that "yesterday" assigns a semantic role and that "lychee" does not, hence its semantic dislocation from the rest of the sentence in (78).

Returning now to the analysis in (80), one question that naturally arises concerns the precise nature of the semantic role which "yesterday" assigns to its syntactic head. In previous sections I showed how it might be possible to derive the more subtle semantic roles which link verbs and their arguments, such as agent, patient, goal and experiencer, from lexical properties of the predicate verb in conjunction with just two, basically contentless primary semantic roles, which I labelled Subject and Object. The minimal assumption would be that these same two primary roles underpin the semantic relations which link adjuncts with their syntactic heads, and this is the possibility that I will explore here. Quite simply what I am suggesting is that an adjunct may assign a Subject role to its syntactic head (although I will modify this claim later):

81)  

Thus, to borrow WG terminology for a moment, "drink" in (81) is the 'tomorrow-er', or that which will occur tomorrow. Similarly, while "in" assigns an Object role to "Bucharest" in (79ii), so the same preposition can then assign a Subject role to "buy":

82)  

Just as semantic properties of the predicate "yesterday" define a temporal content to its Subject role, so analogous properties of "in" determine that its own Subject role will be locational.

It may certainly seem strange initially to think of "drink" as the Subject of "yesterday" and "buy" as the Subject of "in". Bearing in mind, however, that Subject and Object simply refer to two basically contentless semantic roles in LG, there is no reason to think that adjuncts too should not be free to participate in these roles. Indeed, one of the advantages of this approach is that it unifies the otherwise diverse semantic properties of different words. In (82), for example, the preposition "in" and the verb "buy" are both semantic predicates with exactly the same argument structure; both assign Subject and Object roles, and in both cases the
precise content of these roles is determined by the predicate. Similarly, the adverb “yesterday”, in assigning a Subject role only, displays the same semantic structure as an intransitive verb.

Naturally, the semantic properties of words like “yesterday” and “in” are independent of their status as syntactic adjuncts; in section 4.3.6 I suggested that the distinction between arguments and adjuncts may be contextual rather than absolute, and in cases where a head is available to license “yesterday” and “in” actively, as syntactic arguments, the semantic properties of these words will remain the same. For example, in (83) below the preposition “in” is an argument of the verb “put”, but still assigns a Subject and Object role:

83) I

Ted [FIN put] the booze in his shed

Ted put the booze in his shed'.

Here the verb “put” actively licenses two dependents, one of which is the phrase headed by “in”. According to the principle of Subject identification in (65), “the (booze)”, being the nearest d-commanding argument to “in”, will qualify as its Subject, precisely the interpretation associated with (83). In this way it is possible to capture an important distinction in interpretation between (82) and (83); in (82) it is the act of “buying” which is the argument of “in”, that which is “in Bucharest”. In (83), however, it is the object “the booze”, and not the act of “putting” itself, whose situation is described as being “in the shed”. I know of no other theory that is able to offer a principled explanation for this contrast.

However, this last point does raise a residual problem with the representations in (81) and (82); in these examples, unlike (83), the putative Subjects of “yesterday” and “in Bucharest” are their respective syntactic heads, the verbs “drink” and “buy”. Recall from the discussion in section 4.4.3 though that no element can d-command its own direct or indirect dependent. For this reason, then, the verbs “drink” and “buy” in (81) and (82) will not d-command their adjunct dependents, and thus - according to the principle of Subject identification - should not qualify as their Subjects. Evidently, then, assuming that we are not to make an untidy stipulation about the Subjects of adjuncts, the representations in (81) and (82) will have to be modified in some way so as to allow the verbs to d-command their adjuncts. In fact the only possible way in which this can be accomplished is to allow the FIN element, rather than the verb, to license the adjunct in question:
II. Ted [FIN drink] aftershave yesterday

Here, then, in each case the adjunct and verb will be 'sisters' in the sense that they share the same head, FIN. In this way the verb will qualify as the nearest d-commander for the adjunct, and will consequently be assigned its Subject role\(^2\).

Of course one implication of the analysis in (84) is that adjuncts no longer bear any direct syntactic relation to whatever they modify, and it may thus initially appear that the syntax of these structures has been sacrificed for the sake of maintaining a uniform principle of Subject identification. This, however, is not the case; syntactic structure in LG is purely a matter of licensing, and has no direct correlation with interpretation. As I have already pointed out, the basic characteristic of the passive licensing relation, which distinguishes adjuncts from arguments in LG, is that it is sponsored solely by the dependent. In other words the relation between "buy" and "in" in (82), for example, arises solely from the latter's need for a head, and is not registered in any way in the verb. From a purely licensing point of view, then, any head, including FIN, or even "Ted", could license an adjunct passively, just so long as it does not remain unattached to the rest of the sentence. The fact that an adverb serves to 'modify' a verb is a matter of interpretation rather than syntax, and consequently the important link between adjuncts and verbs is semantic. All that is required, then, is that whatever word licenses the adjunct allows for the correct semantic interpretation to be derived from the resulting syntactic structure. What I am suggesting, then, is that in the case of adjuncts, unlike arguments, the desired semantic structure should be allowed to determine the syntactic structure.

\(^2\)Notice, however, that in both examples the Subject "Ted" is also a sister of the verb and the adjunct, since it also shares the same head, FIN. "Ted" will therefore also d-command the adjunct while also being equidistant from it with the verb. Technically, then, both the verb and the subject "Ted" should qualify as potential subjects for the adjunct. I assume that in ambiguous cases like these the predicate will simply assign its Subject role to whichever of the two candidates is the most suitable Subject for it. I will say more about this later.
As I will discuss briefly in the following two sections, there are actually some specific syntactic advantages to allowing FIN rather than the verb to license adverbs. For the moment, however, I hope to have offered here an indication of how a simple system of semantics based on just two primary roles can be extended to adjuncts too; typical adjuncts, such as prepositions or simple adverbials, can be thought of as predicates which assign a Subject and possibly an Object role in exactly the same way as verbs do. Quite apart from bringing a desirable uniformity to semantic structure across a range of syntactic constructions and categories, this proposal has another welcome consequence: the introduction of ‘passive licensing’ in section 4.3.6 introduced an element of unconstrainedness into LG, since any head can passively license virtually anything. Here, then, I hope to have shown how this potential unconstrainedness can be checked by independent principles of semantics; any word that is passively licensed by a head, X, will, by definition, not be assigned a semantic role by X. Consequently the passive licensee, or adjunct, itself must guarantee its own semantic integration with the rest of the sentence. Therefore only those passive licensing relations which allow this semantic integration to take place will be acceptable.

4.4.7 - Word order in LG

One topic that I have not yet addressed in any detail is word order in LG. Following the discussion in section 4.3.4, I assume that the licensing properties of elements stored in the lexicon are entirely non-directional. Thus returning once again to the lexicon fragment in (37) in section 4.3.4 above, the information displayed there relates only to words’ capacity to license (and be licensed by) others, and gives no information as to where an element’s head or dependent should be situated in relation to it; as I pointed out at the time, the ordering of words in a syntactic structure will have to be determined by independent principles which have nothing to do with the lexically-specified licensing properties of individual elements. In this way LG maintains the Immediate Dominance/Linear Precedence distinction first formulated in GPSG (Gazdar and Pullum 1981, Gazdar et al. 1985).

In common with most other syntactic theories, I assume that information concerning word order in LG can be expressed in terms of a head parameter, which states a generalised linear ordering between heads and dependents (Chomsky 1970, 1981, Hudson 1984-a). According to this system of parameterisation a language is predicted to be either head-first, in which case heads will generally precede dependents, or head-last, in which case dependents
usually precede heads. English is almost always classed as a head-first language in that heads
generally precede their dependents (Greenberg 1966, Tomlin 1986). This parameter setting for
English can be stated simply as follows:

85) $X < Y$

This setting alone accounts for a wide variety of word order phenomena in English, including
the ordering of prepositions and determiners before their dependent nouns, and verbs before
their objects. So too (85) accounts for the fact that complementisers precede the subordinate
verbs that they introduce. Furthermore, if the assumptions made in Kreps (1996-a) are correct
and adjectives are analysed as heads of nouns, the head parameter setting in (85) will apply in
this case too.

However, the head-first setting in (85) does not account for all word order phenomena,
and there are certain circumstances in which dependents occur before their heads in English,
notably in the case of subjects, which usually precede their head, adverbs, which are generally
free to occur on either side of their head, and extractees. Putting aside for the moment the
issue of extraction, which will be dealt with extensively in the final chapter, this leaves us with
subjects and adjuncts. Different theories have resorted to various methods to account for the
position of these elements. In PP Theory, for example, subjects are said to occupy a specifier
position, [spec,IP], which, in English, precedes the head of the phrasal projection while
adjuncts occupy an adjoined position which is generally free to occur on either side of the
head. As I described in section 3.3.3 of the third chapter, WG’s solution to this question was
to set up a fundamental distinction between pre-dependents and post-dependents. Subjects,
along with visitors, were then described as instances of pre-dependents while adjuncts could
be classified as instances of either category, as pre-adjuncts or post-adjuncts.

I believe, though, that it is possible to formulate a more straightforward solution in LG,
taking into account some of the assumptions made earlier. More specifically, following the
discussion in sections 4.3.3 and 4.4.6, one thing that both subjects and adverbs have in
common is that they are both licensed by FIN. Indeed, apart from the verb with which it is
fused, these are the only elements which are licensed by FIN, subjects as active licensees
(arguments) and adverbs as passive licensees (adjuncts):
I believe, then, that it might be possible to link the exceptional behaviour of adjuncts and subjects vis-à-vis the head parameter with the fact that these are licensees of the same head, FIN. Quite simply all dependencies will now obey the head-first parameter setting in (85) with the sole exception of those headed by FIN. One possibility would be to supplement (85) with an additional proposition stating that dependents of FIN can occur on either side of it. This more specific statement relating to FIN could then override the more general parameter setting in (85), rather along the lines of the overriding system in WG described in section 3.3.2 of the third chapter. However a possibly simpler approach might be to combine the generalised head parameter setting in (85) with a more specific opt out relating to FIN:

87) - Head Parameter (revised)

\*X < Y (where y ≠ FIN)

Unlike (85), which states a positive specification for head-dependent ordering, (87) instead prohibits dependent-head ordering, except in cases where the head is FIN. The effect of this is to allow dependents of FIN the freedom to occur on either side of their head. So subjects, for example, will generally precede their head, but will follow it in cases of auxiliary inversion. Ultimately it may prove possible to modify (87) further, seeking perhaps to derive FIN's exceptional behaviour from the fact that it constitutes the root of most clauses.

4.4.8 - 'Control' phenomena in adjunct clauses

Quite apart from its useful implications for the formulation of the head parameter, the assumption that FIN is responsible for licensing adverbials can also help account for 'control' phenomena in adjunct clauses. Consider the examples in (88):

88)  
  i. I met Ted [after leaving the clinic]  
  ii. I met Ted [after I left the clinic]  

The two examples here are basically synonymous. In other words, the subject of the verb in
the bracketed adjunct clause in (88i) is interpreted as “I” rather than “Ted”. Consider now (89), the syntactic structure of (88i):

89)  
I [FIN meet] Ted after [ING leave] the clinic

I assume that “after” is a two-place predicate which will assign a Subject role to “meet” as well as an Object role to the gerundive “leaving”. Consequently “after” will have to be passively licensed by FIN so that it will be d-commanded by the verb “meet”, the recipient of its Subject role.

Turning now to the semantic argument structure of “leave”, in seeking its Subject we must locate the closest d-commanding argument. Moving up the dependency chain from “leave”, the first head to license additional dependents is FIN. These dependents, “meet” and “I”, will therefore be the closest d-commanders to “leave”, and will thus both qualify as potential Subjects for it. Assuming that “leave” assigns an agent-type Subject role, though, the most suitable recipient of this will be “I” rather than “meet”:

90)  
I [FIN meet] Ted after [ING leave] the clinic

In this way the assumption that FIN licenses the adjunct clause headed by “after” correctly accounts for the interpretation associated with the example in (88i). If, however, the verb were instead to license the adjunct clause then the intended reading would not be possible:

91)  
*I [FIN meet] Ted after [ING leave] the clinic

Here the closest d-commanding argument to “leave” is “Ted”. Thus the structure in (91) incorrectly predicts that in (88i) “Ted” rather than “I” should be interpreted as the Subject of “leave”. This therefore lends further support to the suggestion that FIN rather than verbs should license adverbs.

However, at the end of section 4.4.6 I suggested that in the case of passive licensing relations syntactic structure could be determined largely by the desired semantic interpretation. The assumption that FIN licenses adjuncts is therefore not a matter of principle, but is instead a reflex of the fact that most adjuncts ‘modify’ (assign a Subject role to) verbs. In cases where
adjuncts do not modify verbs, though, a different licensing structure will be required. Consider now the examples in (92) below:

92) i. [Leaving the clinic] I met Ted.
   ii. I met Ted [leaving the clinic].

In (92i) the only possible Subject of “leave” is “I”. In (92ii), however, the subject of the adjunct clause could be either “I” or “Ted”, depending partly on intonation. As far as I know there has never been a satisfactory, principled explanation for this contrast in any theory. However, taking into account some of the assumptions made previously about the licensing of adverbs as well as the head parameter, I believe that it might be possible to offer a relatively simple account of these data in LG.

The first point to note is that the adjunct clauses in (92) are not headed by a predicate complementiser, such as “after” in (88) or “while”, and consist instead of a simple gerundive and its dependent. One implication of this is that there is no predicate in these adjunct clauses which needs to assign a temporal-type Subject role to the matrix verb. For this reason, unlike (88i)/(89) above, it is not necessarily the case here that FIN is responsible for licensing the adjunct. The only semantic link between the bracketed adjunct clauses in (92) and the matrix clauses will actually be the Subject role assigned by “leave”. From an interpretative point of view, then, there is no reason why either “I” or “Ted” in the matrix clause should not be assigned this Subject role. Bearing in mind, then, that the intended interpretation of an adjunct will control its syntactic relation with its head, (93i) and (ii) below represent the two possible syntactic structures for (92ii):

In (93i), if “leave” is to assign its Subject role to “I”, then the adjunct clause, headed by ING, must be licensed by FIN so as to allow its other dependent, “I”, to d-command “leave”. In (93ii), however, in assigning its Subject role to “Ted”, “leave” must be d-commanded by “Ted”, requiring that the adjunct itself is licensed by the verb “meet”.

171
Turning now to (92i), one might initially assume that structures analogous to those in (93) will apply in this case also, in other words that the adjunct clause could be passively licensed either by FIN or "meet" so as to allow "leave" to assign its Subject role either to "I" (94i) or "Ted" (94ii):

\[
\begin{align*}
94) \text{(i)} \quad & \text{[ING leave] the clinic I [FIN meet] Ted} \\
\text{(ii)} \quad & \text{[ING leave] the clinic I [FIN meet] Ted}
\end{align*}
\]

However, we know that the interpretation associated with (94ii), where "Ted" is taken to be the Subject of "leave", is impossible for the sentence in (92i). We thus have to find a way of ruling out the structure in (94ii) while allowing that in (94i). In fact (94ii) can be excluded fairly easily on the grounds that it violates the revised head parameter setting stated in (87) above. According to this all dependencies must be head-first except for those headed by FIN, whose dependents are free to occur on either side of it. In (94ii), however, the verb "meet" is shown to have a pre-dependent in the form of ING, a violation of (87). (94ii) is thus ungrammatical, and its associated interpretation, with "Ted" as the Subject of "leave", is consequently excluded.

In this way the assumptions outlined above concerning the licensing of adjuncts, in combination with the revised head parameter stated in (87), allow us to account in a fairly straightforward and principled way for the contrast in (92); if the adjunct clause precedes the matrix clause, as in (92i), then it can only be licensed by FIN, and thus the only possible Subject for "leave" is "I". If, though, the adjunct clause follows the main clause, as in (92ii), then it can be licensed either by FIN or the verb, thus allowing both "I" and "Ted" to be possible Subjects for "leave". Returning now to the example in (88i)/(90), the fact that the adjunct clause here is headed by "after", a predicate which assigns a Subject role to the matrix verb, means that "after" itself can only be licensed by FIN, regardless of its position in relation to the matrix clause. For this reason the only possible Subject of "leave" in (88i)/(90) is "I".
4.5 - Conclusions

In this chapter I hope to have shown how it might be possible to formulate a simple but principled theory of dependency syntax based on licensing. This outline theory, which I have referred to as Licensing Grammar, or LG, takes its starting point from Word Grammar (WG), discussed in the previous chapter. Like WG, LG is a lexicalist theory which seeks to reduce as much grammatical knowledge as possible to properties of individual lexical items. So too both WG and LG are monostratal theories in that they recognise only a single level of syntactic structure without seeking recourse to syntactic movement or empty categories.

However, many of the exploratory assumptions underpinning LG have parted company in fundamental ways from what is standardly assumed in WG and other dependency theories. Perhaps the most important of these differences is the explicit recasting of dependency in LG in terms of licensing. In section 4.2.1 I argued that this reformulation was necessary in order to invest the dependency relation with a meaningful and coherent content. In this way dependency can serve more usefully as an autonomous and dynamic tool of syntactic theory in its own right rather than as a blanket term covering a variety of distinct, and more contentful, GRs. Moreover, the additional content that licensing brings to dependency is also valuable in that it makes each syntactic relation more ‘visible’ and thus its existence more open to empirical scrutiny. This points to a more constrained and transparent theory of syntax in which, as I discussed briefly in 4.3.2, the licensing structure of any sentence could be discovered largely on the basis of simple deduction with a minimum of theoretical assumptions.

Another consequence of reformulating dependency in terms of licensing is that it forces us to adopt a radical position on the recognition of GRs; quite simply the licensing relation, in expressing the existential contingency of one word upon another, cannot really be arbitrarily subclassified, and is therefore incompatible with the recognition of distinct, labelled subtypes in the form of GRs. In this way LG is unlike any other DG in that it is ‘monorelational’. In section 4.2.3 I showed how abandoning GRs may have some desirable consequences for LG in reducing the size of the grammar and making the theory more constrained. Moreover, by eschewing GRs LG can claim to be a truly lexicalist theory; since the grammar need not contain information pertaining to distinct relations, the syntax of LG will essentially consist of a head parameter and a lexicon, in which individual words specify how they participate in the unitary syntactic licensing relation. In section 4.3.4 I drew attention to certain parallels between this ‘ultra-lexicalist’ view of dependency syntax and Categorial Grammar.
In section 4.3.3 I introduced the FIN element, an independent sub-lexical entity which comprises information pertaining to finiteness, and possibly tense, aspect and polarity too. I suggested that this FIN element might be responsible for licensing subjects, thus accounting for their absence with non-finite verbs. Of course the putative presence of FIN entails that finite verbs will not enter into any direct syntactic relation with their subject, thus deriving a non-configurational equivalent of the VP. Another implication of recognising sub-lexical elements such as FIN in LG is that the licensing relation might apply equally to the internal morphological structure of words as well as to inter-word syntactic relations. I explored this possibility in section 4.3.5, where I suggested that elements, rather than words, might be the basic constituents of syntax whose licensing properties are stored in the lexicon. Of course individual words will usually comprise a single element, although some words, like finite verbs, genitives and gerundives, may consist of two (or more) fused elements. I argued that it might thus be possible that ‘syntactic’ and ‘morphological’ structure in LG are one and the same thing, with any distinction between the two being a matter of where individual languages superimpose word structure on combinations of elements linked by licensing relations.

The introduction of the FIN element had an other important consequence for LG, in that it introduced a disparity between syntactic and semantic structures; since verbs do not license their subjects, evidently any semantic relation between the two will not be reflected directly in the syntactic structure. I discussed this issue in section 4.4, where I examined the possibility of capturing semantic predicate-argument structure through a simple system of just two primary roles, which I labelled Subject and Object. I suggested that these semantic roles could be derived in a principled way on the basis of syntactic licensing structure, and in sections 4.4.3 - 4.4.5 I examined how an inherently flexible principle of Subject identification, based on the structural relation of d-command, allows semantic Subjects to be located in a uniform way over a wide variety of syntactic constructions. In section 4.4.6 I raised the possibility that the same binary system of semantic roles might also apply to adjuncts, with adverbs and prepositions actually assigning a Subject role to whatever they serve to ‘modify’. Not only does this proposal eliminate the need to recognise a distinct primitive semantic role for each type of adjunct, in sections 4.4.7 and 4.4.8 I suggested that it may also have desirable consequences for the formulation of the head parameter in LG as well as in accounting for a range of complex ‘control’ phenomena in non-finite adjunct clauses.

174
Of course I cannot hope to have offered anything more than an exploratory discussion of LG here. Certainly much work remains to be done if any of the points raised in this chapter are to be examined in any detail. For example the controversial claim that a dependency theory can survive without recognising distinct GRs will have to be tested against a broader range of data before it can be firmly accepted or rejected. So too the simple LG syntax I have described as consisting entirely of a lexicon and a head parameter setting will almost certainly have to be supplemented with additional principles and mechanisms in order to account for more complex data such as passivisation and extraction (the topic of the next chapter). Hopefully, though, at least some aspects of the outline theory sketched in this chapter are of some intrinsic interest and value, and can stand on their own as proposals which might merit further exploration. However the main relevance of LG here lies in its potential to offer an effective and plausible dependency-based account for the complex extraction data reviewed in the first three chapters. This will be the subject of the fifth and final chapter.
CHAPTER 5
LICENSING GRAMMAR, PROCESSING AND EXTRACTION

5.1- Introduction

In the previous chapter I outlined a model of dependency syntax, Licensing Grammar (LG), a theory which was shown to differ substantially from other dependency-based frameworks, notably in its failure to recognise distinct GRs and in its assumed disparity between syntactic and semantic structure. In this final chapter I wish to return to the central topic of this thesis and explore how the theory of LG might serve as the basis of a holistic and principled account for the extraction data examined in the first three chapters. However, as I suggested at the end of chapter 3, I believe that a purely structural syntactic account of extraction may be inappropriate, and I will argue here that in order to capture many of the intricacies of the data we must look to issues of language processing as well as syntax. One of the main concerns of this chapter will thus be to show how the syntactic mechanisms of LG underpinning extraction can be conditioned and constrained by the operation of the parser. I believe that only in this way is it possible to account for the scalar nature of grammaticality judgements associated with extraction phenomena, as well as some of the pragmatic influences which affect the data.

I have suggested that extraction data can be viewed as a sort of benchmark test against which the adequacy of competing theories may be evaluated. My aim here, then, is to demonstrate that a relatively simple, though unconventional, dependency-based theory of syntax such as LG, when supplemented with processing factors, can offer a plausible and effective account of these data, and one which in my opinion goes beyond anything formulated in other dependency frameworks. In this way I hope to demonstrate for the first time that a dependency theory might after all be able to compete on an equal footing with PP Theory in its treatment of extraction.

In section 5.2 I will outline some key issues in language processing and explain in greater detail why I believe that these are relevant to extraction data. Section 5.3 will then describe a simple LG-based parsing system; I will show, among other things, how the operation of this parser may derive the adjacency constraint which was noticeably absent in LG itself. Section 5.4 introduces LG's syntactic approach to extraction, and demonstrates how this
can be constrained by the operation of the parser. Finally in sections 5.5 and 5.6 I will show how island constraints as well as extraction asymmetries can be explained in terms of processing difficulty.

5.2 - Some Processing-Related Issues

5.2.1 - The need for processing

LG, like WG, is a monostratal dependency theory, and thus any account of extraction advanced within this model will have to involve a single surface structure of binary relations alone. Unlike WG, however, LG is a ‘mono-relational’ theory which does not recognise distinct GR’s, and thus will evidently not be able to invoke a separate relation such as WG’s Visitor in order to help account for the long-distance dependencies arising from extraction. From this point of view, then, LG is doubly constrained in that in seeking to account for extraction phenomena the theory can only call upon a single level of syntactic representation composed of a single relation. Consider, for example, the sentence in (1):

1) What did Wally wear?

Given the more constrained principles of licensing upon which LG is based, the only possible structure for (1) will be (2):

2) What [FIN do] Wally wear

Assuming that “what” is a syntactic (and semantic) argument of “wear”, this will be the only word capable of licensing “what”; no other word in (2) can actively license it since none of them requires any other dependent. Similarly, “what” cannot be passively licensed by another word in (2) for the simple reason that it already has a head in the form of “wear”; as I described in section 4.3.6, a passive (dependent-sponsored) licensing relation occurs in cases where a dependent requires a head, but the head does not require a dependent. This is obviously inapplicable in the case of “what”, which is actively licensed by “wear”.

I have already described in the third chapter how analyses like the one in (2) are inadequate. For one thing they offer no explanation as to why the WH-object of “wear” occurs so far from its head at the front of the sentence. (2) also violates well-motivated principles of locality, in that “wear” and its dependent “what” are separated by a number of words which bear no direct relation to them. Moreover, this analysis is totally unconstrained, and there is
nothing in (2) to prevent the extraction dependency from spanning an 'island' domain such as (3):

```
3) *what [FIN do] Wally arrive while he [FIN be] [ING wear]

*‘What did Wally arrive while he was wearing?’
```

In chapters 2 and 3 I described how PP Theory and WG both overcome problems of this sort by reducing the long distance dependencies between heads and extractees to a series of smaller local ones, either by a system of intermediate traces in the case of PP Theory or by the use of ancillary visitor dependencies in the case of WG. The question, then, is whether a similar localisation of long distance dependencies can also be achieved in LG, given that the theory incorporates neither transformational mechanisms nor allows for the existence of additional dependencies or distinct GRs.

I believe that the key to this question lies in processing. More specifically, I believe that although LG provides the syntactic means of coping with extraction, this can only be properly understood by examining how syntactic structures are processed. Very simply, I will argue that LG may allow the licensing capacity of one word to be transferred to another. Thus FIN in (2) above, for example, could license “what” through a valency specification transferred from “wear”. I will describe this system of valency transferal (VT) more fully in section 5.4. My claim will be, however, that the syntactic process of VT is executed ‘online’ by the parser. Furthermore, I will argue that the parser can only carry out each application of VT locally, (in a sense to be defined later). I believe that this procedural approach to syntax may thus allow a series of temporary dependencies to be established between an extractee and its head, thus localising the distance between them.

Relating extraction to processing in this way has a number of advantages; for one thing it raises the possibility that ‘island’ violations could be explained in terms of processing difficulty rather than ungrammaticality. This in turn implies a less absolute and more relative approach to these data; after all, some things can be ‘more difficult’ than others, and this more flexible view seems better-suited to the often scalar nature of island violations discussed in sections 2.5 and 3.5; one of the reasons why I believe that a purely syntactic account of extraction is inappropriate is that it implies an absolute distinction between what is grammatical and ungrammatical. Furthermore, I will also argue that incorporating processing-related factors into the LG account of extraction offers us a realistic hope of taking into account some of the
pragmatic influences on the data, discussed in the first chapter, which would otherwise be very difficult to accommodate within a purely syntactic framework.

The rest of this chapter will be devoted to exploring how simple processing factors may relate to syntax in LG, and how the two may be shown to interact in offering a plausible account for extraction phenomena. I should stress at this point, however, that by this I am not seeking to mask any shortcomings of LG by appealing to processing. Nor is it my intention here to put forward a rigorous, fully-defined or necessarily implementable model of parsing; the main emphasis of this chapter is syntax and how it can be constrained and conditioned by processing considerations, and for this reason I will try to keep parsing-related technicalities and formalism to a minimum.

5.2.2 - A brief overview of parsing

I assume, relatively innocuously I think, that the core meaning of a sentence is a product of its constituent words and the relations between these words\(^1\). The latter point is illustrated by (4) below where two sentences containing the same words display different meanings:

4)

ii. Mao kissed Nixon.

In section 4.4 of the previous chapter I described how in LG the semantic relations (or roles) of a sentence, those of most immediate relevance to its meaning, can be read off or derived from a syntactic structure in a systematic way. This is more or less true of all syntactic theories, and thus the implication is that in order to understand a sentence two key tasks of the hearer will be to recognise its constituent words and to work out the syntactic relations which exist between them. This process of recognition and structure building has traditionally been described as parsing - see Winograd (1983 ch. 3), De Roeck (1983), Jackendoff (1987 ch. 6), Altmann (1989) and Garman (1990 ch. 6) for some general introductory comments.

The basic function of the parser is to assign a well-formed syntactic structure to a grammatical string of words. Thus in some sense the parser can be viewed as a machine which inputs strings of words or elements and outputs syntactic structures created on the basis of

\(^1\)By core meaning I intend to exclude pragmatic components such as implicatures and explicatures (Sperber and Wilson 1986).
these strings. In order to achieve this the parser must, of course, make use of grammatical knowledge, and the parsing process is often described as the procedural implementation of principles and facts stored in the grammar (Marcus 1980, Winograd 1983, Hudson 1990 ch. 2). For example, a PP Theory-based parser might utilise principles of X' Theory as well as Theta Theory or Case Theory in order to construct a phrase structure tree such as (5ii) on the basis of the input string in (5i) (Berwick and Weinberg 1984, Berwick 1991, Pritchett 1992):

5) i. A B C ii.

\[
\begin{array}{c}
\text{CP} \\
\text{AP} \\
A' \\
\text{BP} \\
A \\
B \\
C
\end{array}
\]

One important property of the parser is that it must operate incrementally from left-to-right (Marlsen-Wilson 1973, Marlsen-Wilson and Tyler 1980). Theoretically it is of course possible to devise a parser which starts building structure at the end or in the middle of a sentence. Humans evidently do not operate in this way, however, and instead processing gets underway immediately as soon as words are recognised. This, for example, explains the occurrence of 'garden path' sentences where the processor is fooled into building a structure which subsequently proves to be incorrect (Marcus 1980, Frazier and Rayner 1982, Pritchett 1992). So, for example, in (5) above an incremental parser might seek to link A and B, creating the partial structure in (6), before C even occurs:

6)

\[
\begin{array}{c}
\text{AP} \\
A' \\
\text{BP} \\
A \\
B
\end{array}
\]

When C occurs its own projection will be added to (6), yielding the possible structure shown in (5ii). The parsing process continues in this way with structure being built up as new words are recognised in the input string.\(^2\)

\(^2\)One issue which pertains particularly to phrase structure parsers is the question of whether the processor operates 'top-down' or 'bottom-up'. Very simply, a 'top-down' (or hypothesis-driven) parser will start by assuming the existence of a top node, usually a clausal constituent such as S' or CP, and will then seek to accommodate words from the input string within the substructures of this higher node. A 'bottom-up' parser, however, will instead build up lower nodes in the phrase structure.
Since the linguistic mainstream has been dominated by constituency-based theories of syntax, it is not surprising that the majority of parsing models have tended to concentrate on how hearers build up constituent structure on the basis of words, illustrated in (5) above; as Gorrell (1995 p. 43) concludes: "It is a rare processing model that does not make some reference to syntactic constituents and structural relations" (see also Winograd 1983 chs. 3 and 7 and Fraser 1993 ch. 1). The first serious attempt to offer a coherent and explicit model of phrase structure parsing can be found in Bever (1970) and Fodor, Bever and Garrett (1974). Their main parsing strategy, the Canonical Sentoid Strategy was to identify sequences of NP-V-NP in the incoming string and to analyse these sequences automatically as subject, verb and object. This is an early example of a parsing axiom or heuristic, an overall strategy which guides the processor in its attempt to build structure on the basis of the input string. Since the work of Fodor et al. (1974) a number of different parsing heuristics have been proposed which seek to build up constituent structure in a simple and principled way, see for example, Kimball (1973), Frazier and Fodor (1978), Marcus (1980), Frazier and Rayner (1982) Berwick and Weinberg (1984) and Pritchett (1992).

Dependency-based parsers, by contrast, have been comparatively rare; Fraser (1993) contains a valuable survey. Nevertheless, parsers have been formalised on the basis of some dependency theories such as WG (Hudson 1989, 1994, Fraser 1989, 1993) and Lexicase (Starosta and Nomura 1986). In addition, various systems of machine-translation and speech recognition have employed a dependency formalism (Hays 1966). Faced with the same string in (5i) a dependency-based parser might output the structure in (7):

\[ A \quad B \quad C \]

Like its constituency-based counterparts, a dependency parser must also operate incrementally from left to right, and will thus start building structure on the basis of words as they are recognised in the input string. However the absence of phrase structure in dependency grammars means that for the most part a dependency-based parser will rely less on structural algorithms and concentrate more on the properties of individual words as they are encountered. Thus the structure in (7) is largely a product of the fact that A and C license tree on the basis of words as they are encountered in the input string, and then seek to create higher nodes by combining these lower ones.
dependents, whereas B does not, and it is valency information of this sort that the parser will have to access in order to construct a well-formed syntactic representation. In this way dependency parsers will, for the most part, be more data-driven than hypothesis driven, and for this reason the top-down/bottom up distinction which characterises phrase structure parsers is of questionable relevance to dependency-oriented processing, although see Fraser (1993).

One issue which pertains equally to dependency and phrase structure parsers is the question of ambiguity. It is often the case that a legitimate input string can be associated with more than one well-formed structure. This is true, for example, in cases such as (8):

8) Visiting relatives can be a nuisance.

In addition to examples of global ambiguity such as these, the parser will also have to resolve temporary (or local) ambiguities such as (9):

9) i. I know the Bishop.
   ii. I know the Bishop likes football.

Both (9i) and (ii) display temporary ambiguity in that at the point when "the Bishop" occurs a purely incremental processor will not know whether it is the object of "know", as in (9i), or the subject of a complement clause, as in (9ii). In order to resolve this ambiguity and establish the correct structure the parser will have to wait and see whether there is any further structure to process.

Following Fodor et al. (1974) it is possible to distinguish two broad approaches to the question of ambiguity, serialism and parallelism; very basically when faced with an ambiguous input string, a parallel parser will construct all possible structures that can be formed on the basis of that string. If then later the ambiguity is resolved, then the parser can go back and eliminate any incorrect structures that it might have built. Presented with the same ambiguous input string a serial parser will construct just one structure. It may of course transpire later that the parser's choice was wrong - this, of course, is one of the hazards of operating incrementally. In this case the parser will have to go back and undo the structure it has created and build another one. This process is called backtracking.

In different ways the creation of parallel structures and backtracking are both costly in terms of processing effort, and Marcus (1980) seeks to eliminate the need for these strategies by developing a model of deterministic parsing. A deterministic parser, like a serial parser, creates only one structure at a time, but uses a look-ahead device or 'buffer' to ensure
that each structural choice it makes is correct. Thus in (9i), for example, before linking “the Bishop” as an object of “know”, a deterministic parser will check to see if this analysis is compatible with subsequent words in the input string, which it clearly is not in the case of (9ii). In this way the parser can avoid the need to backtrack later. Since the work of Marcus (1980) most phrase structure parsers have been either strictly deterministic, in that they allow no backtracking, or weakly deterministic, which may allow only very limited backtracking (van de Koot 1990). So too the majority of recent dependency parsers have been serial and/or weakly deterministic; this is true, for example, of Starosta and Nomura’s (1986) Lexicase parser and Fraser’s (1989, 1993) WG-based parser as well as other systems developed by Hays (1961) and Covington (1990). However, Hudson’s (1989) WG-based parallel parser is an exception.

5.2.3 - Other processing-related accounts of extraction

Even before serious parsing systems were first formulated, people sought to attribute certain linguistic phenomena to processing factors. For example, Miller and Chomsky (1963) and Chomsky (1965) note how a grammar can be simplified if centre-embedded constructions are excluded by processing rather than grammatical criteria. Subsequently a great deal of interest was shown in interpreting aspects of the grammar in processing terms, or vice versa. One of the best-known examples of this was the Derivational Theory of Complexity or DTC (Fodor et al. 1974) which, very simply, saw the parser as a transformational grammar ‘run backwards’. Since then people have generally taken a more autonomous view of the parser and grammar (Abney 1988, van de Koot 1990, Berwick 1991). Nevertheless, more recently, Hawkins (1994) suggests that certain aspects of the grammar may in fact be ‘grammaticalised’ reflexes of processing considerations (see also Ninio 1993-b).

As far as WH-extraction is concerned, there have been a few attempts to explain some aspects of the data in terms of processing considerations. For example Marcus (1980) claims that the presence of the Subjacency constraint in the grammar can be shown to be motivated by the structure and operation of his deterministic parser, a proposal extended and modified by Berwick and Weinberg (1984). Similarly van de Koot (1987, 1990) invokes properties of a Barriers-based deterministic parser to explain the structural relation of L-marking, which is an important component of Chomsky’s (1986-a) approach to WH-extraction, as discussed in section 2.5 of the second chapter. Rather than seeking to find a functional motivation for
grammatical principles in the parser, Pritchett (1991) goes further and suggests that at least some island violations can be excluded by direct reference to the parser's operation. He describes how structures involving violations of certain islands can be shown to confound his central parsing heuristic, the Theta Reanalysis Constraint or TRC, the implication of this of course being that these island violations can be classed as being unprocessable rather than ungrammatical. This is an interesting and radical proposal, and one which raises many issues which I will explore further in this chapter.

However, in spite of the intrinsic interest of Pritchett's theory, I will not discuss it further here, nor any of the other proposals mentioned above, for the simple reason that without exception they have all been formulated on the basis of some form of PP Theory-oriented parser, which inevitably raises numerous issues which are tangential to the central concerns of this chapter. For example, all of the parsing systems mentioned above are phrase structure-based and utilise components specific to PP Theory such as X-bar theory and Theta Theory. Moreover, one crucially important job of a PP Theory-based parser in processing extraction structures will be to locate traces (Stowe 1986, Kurtzman et al. 1991), something that a monostratal dependency-based parser will not have to do.

The main point I wish to stress here is that tentative attempts have been made to explore certain aspects of extraction from a processing point of view, although it is fair to say that on the whole these have been relatively few and far between compared with purely syntactic treatments of the same phenomena. However, as far as I am aware, no such processing-related approach to extraction has ever been proposed within the context of dependency theory. This is perhaps not altogether surprising given the relatively underexplored nature of dependency-based parsing generally. What follows then is, I believe, the first serious attempt to put forward a coherent account of extraction phenomena formulated on the basis of a dependency grammar and its associated parsing strategy.

---

3Pickering and Barry (1991) present some psycholinguistic evidence suggesting that from a parsing point of view a dependency structure without traces is preferable to a phrase structure representation. Moreover, as discussed in Kreps (1994), even psycholinguistic evidence which purports to support the existence of traces such as Stowe (1986), Swinney et al. (1988) and Nicol 1989) can in fact also be shown to support equally a traceless analysis involving intermediate dependencies, such as Hudson's visitor.
5.3 - LG and Parsing

5.3.1 - Some preliminary considerations

In the previous chapter I described LG as an ‘ultra-lexicalist’ theory and suggested that its syntactic components might consist almost entirely of a lexicon and a simple head parameter setting. This simplification of the grammar was shown to be achieved at least in part as a result of the putative absence of distinct GRs in the theory. If this is true then there is no reason why virtually all syntactic information in LG cannot be stored in the form of individual words’ lexical requirements. As I described in section 4.3.1, the syntactic information of a word’s lexical entry will have to include, at the very least, its licensing capacity along with any categorial restriction imposed on its dependent(s).

This relatively simple view of syntax has some important implications for processing; on recognising a word, an LG parser (LGP) will simultaneously access the licensing properties of that word. Structure-building can then proceed directly on the basis of these licensing requirements. In one respect at least, then, the task of an LGP will be simpler than that of other dependency parsers, in that it will not have to work out what sort of relation exists between words. As I noted in the previous chapter, the assumption that there is only one syntactic licensing relation means that words’ participation in this relation is equivalent to a simple binary combinatorial requirement, reminiscent of Categorial Grammar. Essentially, then, the task of an LGP will consist of recognising words, accessing their combinatorial requirements, and satisfying them. In this sense parsing can be seen as a passive rather than active operation which simply responds to words’ licensing properties.

The LGP which I will explore here shares many features common to other dependency-based parsers, notably in the fact that it is incremental and ‘bottom up’, since it operates almost entirely on the basis of words’ individual licensing requirements as they are encountered. In addition, like most other contemporary parsing systems, the LGP is also serial; I assume that when faced with an ambiguous input the parser does not build parallel structures, but instead

\[
\text{THIS}^N \quad X^N
\]

The head parameter will specify independently that the dependent will be found to the right of “this”.

In this way a word’s syntactic requirement for a dependent is, in effect, the same as an instruction to the processor to search for a suitable dependent (Ninio 1993-a). Thus the lexical entry for “this” below could be interpreted as an instruction telling the processor to find a nominal dependent:

4

185
seeks to create a single structure. Processing errors which arise from this mode of operation are remedied by very limited backtracking, whose application, as I will show in section 5.3.5, is severely constrained by the very nature of structures created by the processor.

However the LGP to be explored here also embodies various important differences from other parsers, the significance of which should become clear later. Unlike most other dependency parsers, for example, the LGP employs a pair-based rather than a search-based control strategy (Fraser 1993 ch. 12). By this I mean that in seeking to build licensing structure the LGP selects pairs of adjacent words from the input string and then seeks to establish dependencies on the basis of the licensing properties of both of these words. This is unusual since most dependency parsers are search-based in that they operate by accessing a single word’s valency properties at a time and then search for another word in the input string which satisfies these requirements.

Another notable feature of the LGP which sets it apart from most, though not all, other dependency parsers is that it makes use of a rudimentary system of derived constituency. As I will describe in greater detail later, when the parser forms a dependency between two words, it creates a cell around them. Thus once A and B have been linked by a licensing relation they will constitute a cell (A B). If then another word C can be related either to A or B then a new cell will be created around (A B) and C: ((A B) C). In this way cell structure will be built up incrementally reflecting the order in which different dependencies have been established by the processor. The resulting derived constituency has no relevance other than as a processing history of the structure, and I assume that a cell will have no independent properties of its own over and above those derived from the words which constitute it.

5.3.2 - The structure of the processor

The LGP which I will explore here is extremely simple, and consists of just three basic components: a list, a stack and a single combinatorial operation, which I will refer to as Form-Dependency or Form-D. These three components are explained in greater detail below:

---

5Fraser’s (1989, 1993) parser also utilises primitive processing-derived constituents known as ‘molecules’. However, the nature and function of these molecules are very different from those of the cells that I am suggesting here.

6My usage of the terms ‘list’ and ‘stack’ is somewhat unconventional, and differs from the way in which these terms are used in more conventional ‘shift-reduce’-parsers (Hudson 1994).
i - The list is a two-place holding area where two adjacent words from the input string are placed while their licensing properties are examined in order to determine whether a dependency can be formed between them. The list is represented by a line, with each of the two places denoted by a position, P1 and P2:

\[
\begin{array}{cc}
& \text{P1} & \text{P2} \\
\end{array}
\]

The two positions of the list P1 and P2 may hold either a single word or a cell, a cell being a unit which comprises two (or more) words which have been linked by a dependency relation. Initially the first word to be recognised in the input string will be placed in P1 and the second in P2. As the parser works through the incoming string, new words are placed in P2 while dependency structure (and corresponding cell structure) is built up in P1. The parsing process will be completed when a well-formed structure is contained in P1 and P2 remains empty.

ii - Form-D is the central processing operation which drives the LGP. Its function is to form a dependency between the two words placed in P1 and P2 of the list. To achieve this Form-D simply accesses the dependency requirements of these words and responds to them. Say, for example that the first two words recognised in a string are W1 and W2. These will be placed in P1 and P2 respectively:

\[
\begin{array}{cc}
W1 & W2 \\
\text{P1} & \text{P2} \\
\end{array}
\]

Here W1 licenses a dependent and W2 requires a head; if W2 conforms to any categorial restrictions imposed by W1 on its dependent, then Form-D will establish a dependency between the two words. In cases such as this where Form-D operates successfully, a cell will be created around the two words. This single cell will then occupy P1 of the list:

\[
\begin{array}{cc}
(W1 W2) & - \\
\text{P1} & \text{P2} \\
\end{array}
\]

This then leaves P2 free to accommodate the next word of the input string to be recognised:

\[
\begin{array}{cc}
(W1 W2) & W3 \\
\text{P1} & \text{P2} \\
\end{array}
\]
Form-D can now apply once again and seek to create a dependency relation between W3 in P2 and either of the two words within the cell in P1. This process will continue until all incoming words have been incorporated into a cell structure in P1 and P2 is empty. Given the head parameter setting for English, Form-D will generally link heads in P1 with their dependents in P2, although an exception will be made for licensees of FIN. I will say more about this later.

iii - The Stack is a simple first-in, last-out holding structure where words are placed when the operation of Form-D is unsuccessful and no relation can be established between the words in P1 and P2. I assume that the stack is a sub-part of the list, and is located in P1. Thus returning to (11) above, if for any reason Form-D cannot establish a dependency between W1 and W2, then W2 will be placed on top of W1 in the stack. I will use a backslash (\) to denote the position of words in the stack, the word occurring to the right of the backslash being placed on top of the word occurring to the left:

\[ \begin{array}{c}
W1 \backslash W2 \\
PI & P2
\end{array} \]

Once W2 is placed above W1 in the stack it alone will remain visible for any new application of Form-D. Thus if a new word W3 is placed in P2 then Form-D will only be able to link this word with W2; W1 will remain 'invisible' until such times as W2 is removed from on top of it. This illustrates the 'first-in, last-out' nature of the stack. If, though, a relation can be established between W2 and W3, then W2 will be removed from the stack and placed back in P2 where a new cell comprising W2 and W3 will be created:

\[ \begin{array}{c}
W1 (W2 W3) \\
PI & P2
\end{array} \]

I assume that removal of a word from the top of the stack - or 'popping' as it is sometimes known - will be the only way in which a cell can be created in P2 rather than P1. As far as (15) is concerned, Form-D can now apply once again in seeking to create a dependency relation between W1 in P1 and W3 in the cell in P2 (Form-D has, of course, already tried and failed to link W1 and W2).

From the above discussion it should be clear that the operation of Form-D can have two possible outcomes; it will either be successful and establish a dependency between the
contents of P1 and P2 of the list, or it will be unsuccessful and fail to establish any relation between them. In the former case a cell is created which occupies P1 of the list, in the latter case whatever occupies P2 will be placed above the contents of P1 in the stack. Note then that in both cases the result is to move a word from P2 into P1, thus leaving P2 free to accommodate a new word from the input string.

Perhaps the easiest way to see how the three basic components of the parser function is to work through the possible parsing permutations of a simple string, A-B-C. The first two elements to be recognised are A and B. These will thus be placed on P1 and P2 of the list respectively:

16) \[ \begin{array}{c}
A \\
P1 \\
P2 
\end{array} \begin{array}{c}
B \\
\end{array} \]

Form-D will now operate and seek to establish a syntactic dependency between A and B. There are two possible outcomes; if the licensing requirements of A and B are suitable, Form-D will succeed, thus creating a cell \((A \ B)\) which will occupy P1:

17) \[ \begin{array}{c}
(A \ B) \\
P1 \\
P2 
\end{array} \]

Alternatively, the licensing properties of A and B will not allow a syntactic relation to be established between them, and Form-D will fail. B will then be placed on top of A in P1 in the stack:

18) \[ \begin{array}{c}
A \ \neg \ B \\
P1 \\
P2 
\end{array} \]

Let’s assume for the moment that Form-D had succeeded and that a dependency relation had been established between A and B, as illustrated in (17). The next word to be recognised in the string is C. This is placed on P2 of the list:

19) \[ \begin{array}{c}
(A \ B) \\
P1 \\
P2 
\end{array} \begin{array}{c}
C \\
\end{array} \]

Once again there are two possibilities here; Form-D might establish a relation between C and one of the elements in the cell \((A \ B)\), resulting in a new cell \((A \ B \ C)\) in P1. Alternatively, Form-D will fail to establish a relation between C and A or B, and thus C will have to be placed above \((A \ B)\) in the stack in P1. These possibilities are illustrated in (20i) and (ii):
Now returning to the example in (16) above, let's assume that Form-D has originally failed to establish a relation between A and B, as illustrated in (18). Once again, the next word to be recognised in the input string is C:

21) \[ \begin{array}{c}
A \backslash B \backslash C \\
P1 & P2
\end{array} \]

Here Form-D will try to establish a dependency between B, at the top of the stack in P1, and C in P2. A is 'invisible' for the time being since it is underneath B in the stack, and will only become visible to Form-D once B has been removed from above it. If the licensing requirements of B and C allow Form-D to operate successfully, then B will be removed from the stack and a cell \((B \backslash C)\) will be formed which will occupy P2:

22) \[ \begin{array}{c}
A (B \backslash C) \\
P1 & P2
\end{array} \]

Now Form-D can seek to link A in P1 with the cell in P2. Evidently Form-D has already failed to link A and B. However, if a dependency can be established between A and C then a new cell can be created in P1:

23) \[ \begin{array}{c}
(A \backslash (B \backslash C)) \\
P1 & P2
\end{array} \]

Otherwise the cell \((B \backslash C)\) will have to be placed on top of the stack in P1 above A:

24) \[ \begin{array}{c}
A \backslash (B \backslash C) \\
P1 & P2
\end{array} \]

Referring back to (21) now, if no relation had been established between B and C then C would be placed on top of the stack, above both B and A:

25) \[ \begin{array}{c}
A \backslash B \backslash C \\
P1 & P2
\end{array} \]

In order to create a well-formed structure from (25) each of the three elements would have to be removed from the stack and incorporated into a single cell structure.
Taking (20), (23), (24) and (25) together now, we can see that on the basis of the three-element string A-B-C, depending on the licensing properties of these elements, the parser can create any of the following five permutations:

26) i. ii. iii. iv. v.
   
   \((A \ B \ C)\)  \((A \ B) \ C\)  \((A \ (B \ C))\)  \((A \ \ (B \ C))\)  \((A \ \ B \ \ C)\)

Of these five possibilities, however, only (26i) and (iii) are viable structures since only these constitute single cells. The permutations in (26ii), (iv) and (v) will have to be processed further with other material in order to remove elements from the stack and create single structures.

5.3.3 - The operation of the LGP

Consider now how the structure in (27) might be processed:

27) The critics [FIN despise] Nyman
   ‘The critics despise Nyman’.

The first two words to be encountered and recognised are “the” and “critics”. These will be placed in P1 and P2 of the list respectively:

28)  
   The critics
   P1 P2

Since “the” licenses a (nominal) dependent and “critics” is a nominal requiring a head, Form-D can successfully operate on the basis of these requirements and establish a licensing relation between the two words:

29)  
   (The critics)
   P1 P2

The next word to be recognised is “despise”, a word composed of two elements, both of which require a dependent:

30)  
   (the critics) [FIN despise]
   P1 P2
Evidently the contents of P1 and P2 in (30) can be linked by a dependency relation since “the” needs a head and both elements of “despise”, FIN and the verb “despise”, require a dependent. Initially, then, it may appear that we are faced with a conflict concerning which of these two elements will license “the”. In fact to resolve this question we need only look to the revised head parameter discussed in the previous chapter and repeated in (31) below:

31)  
\*X < Y (Y \neq \text{FIN})

This states that all dependencies must be head-first in English, except those headed by FIN. Since no actual specification is made for FIN in (31) we can infer that its dependents might occur on either side of it. From a parsing point of view the implication of this is that FIN will always be responsible for licensing any headless constituent which precedes it. As I will show later in section 5.4, this applies equally to extractees as well as subjects.

Returning now to (30), the head parameter guarantees that FIN, and not “despise”, will be the head of “the”, and this will therefore be the only possible application of Form-D:

32)  
\((\text{the critics}) [\text{FIN despise}]\)  
P1  P2

The next word to be recognised, “Nyman”, will be placed in P2 of the list (33i). This can be linked to the existing structure in P1 because “despise” licenses a dependent while “Nyman” itself needs a head (33ii):

33)  
i.  
\((\text{the critics}) [\text{FIN despise}]\)  
P1  Nyman  P2

ii.  
\(((\text{the critics}) [\text{FIN despise}]\text{ Nyman})\)  
P1  P2

In (33ii) a successful parse has been completed since we are left with a single well-formed structure in P1 while P2 remains empty. Now that all the syntactic relations have been established, the axioms of Subject and Object identification discussed in the previous chapter can apply in deriving a semantic interpretation for the sentence.

Now consider the more complex sentence in (34):
34) The critics know that Nyman wrote MGV.

I assume that the first four words, “the critics know that”, will be processed in the same way as the sentence in (27)/(33ii) above, resulting in the structure shown in (35):

\[ (((\text{the critics}) \ [\text{FIN know}]) \ [\text{that}]) \ P2 \]

The next word to be recognised here and placed in P2 is “Nyman”. The only word in P1 which could potentially license “Nyman” is “that”, which in (35) has a spare licensing capacity. Recall from section 4.3.1 of the previous chapter, however, that “that”, being a complementiser, was assumed to license only the finiteness element FIN. Therefore “Nyman”, a nominal, will not qualify as a suitable dependent for “that”. Form-D will thus fail to operate, and “Nyman” will have to be placed in the stack above the cell already occupying P1:

\[ (((\text{the critics}) \ [\text{FIN know}]) \ [\text{that}] \ Nyman) \ PI \]

The next word to be placed in the vacated P2 position is “wrote”:

\[ (((\text{the critics}) \ [\text{FIN know}]) \ [\text{that}] \ Nyman \ [\text{FIN write}] ) \ PI \]

The only word in P1 which is visible to Form-D here is “Nyman”, at the top of the stack. This can of course be linked as a pre-dependent of the FIN element of “wrote” in P2. As a result “Nyman” is removed from the top of the stack in P1 and incorporated into a new cell in P2:

\[ (((\text{the critics}) \ [\text{FIN know}]) \ [\text{that}] \ Nyman \ [\text{FIN write}] ) \ P2 \]

Now the newly created cell in P2, headed by FIN, qualifies as a suitable dependent for the complementiser “that” in P1. Form-D can thus attach FIN as a dependent of “that”, creating a new larger cell in P1:

\[ (((\text{the critics}) \ [\text{FIN know}]) \ [\text{that}] \ (Nyman \ [\text{FIN write}] )) \ PI \]

\[ (((\text{the critics}) \ [\text{FIN know}]) \ [\text{that}] \ (Nyman \ [\text{FIN write}] )) \ P2 \]
The last ‘word’ to be recognised is “MGV”. This will be placed in P2, and then linked with the verb “write” which in (39) still has an unfulfilled licensing capacity:

Once again, then, we are left here with a complete and well-formed structure in P1 while nothing remains to be processed in P2.

It is worth considering one more structure, this time involving an adjunct:

41) Nyman attacked the critics with a mallet.

In sections 4.3.6 and 4.4.6 of the previous chapter I suggested that adjuncts like “with a mallet” are passively licensed, and that their syntactic head is determined by their semantic requirements. So, for example, in (41) the adjunct modifies the verb “attack”; I therefore assume that the head of this adjunct phrase, “with”, will assign a Subject role to “attack”. Recall from section 4.4.3 that Subject roles are assigned to the closest d-commanding argument to the assigner. Thus in order to be d-commanded by “attack”, “with” will have to be licensed by FIN.

42) Nyman [FIN attack] the critics with a mallet

However, from a processing point of view the structure in (41) displays a local structural ambiguity; at the point when “with” occurs there is no guarantee that it will be attached as a passive dependent of FIN so as to modify “attack”. An adjunct phrase headed by the same preposition, “with”, might instead modify the nominal “the critics”, as in (43):

43) Nyman attacked the critics with hats.

Here, then, the verb “attack” will itself have to license the adjunct so as to allow “with” to be d-commanded by “the (critics)”. Similar local ambiguities arise with a wide range of adjuncts:

44) i. Noam hid the banjo from Belgium.
   ii. Noam hid the banjo from Ray.
45) i. Santa broke open the vault in the bank.
   ii. Santa broke open the vault in order to steal the money.
The implication of these examples is that at the point when the adjunct head occurs the parser will lack sufficient information to attach it to anything else in the main clause. One possibility, of course, would be to make an attachment and then backtrack if this later turned out to be wrong, and in section 5.3.5 I will show how this may indeed be a viable strategy in some cases of structural ambiguity. However, I do not believe that backtracking is an appropriate solution here; as I have already explained, the licensing structure of adjuncts is determined by semantic requirements rather than syntactic properties, and there would therefore be no point in making any syntactic attachment, even if it subsequently turns out to be correct, until it is properly motivated by semantics. Moreover, it seems fairly clear that the semantic relation of an adjunct to a clause can only be determined once its own dependent has been located. For example, we only know that "from" in (44i) modifies "the banjo" rather than "hid" once we have linked "from" to "Belgium". What I am suggesting, then, is that a 'transitive' adjunct like "with" or "from" can only be linked to its syntactic head once it has been linked to its own dependent (see also Marcus 1980 ch. 3).

Returning to the structure in (41)/(42), then, I assume that once the main clause has been processed, the next word "with" cannot yet be linked to any word of this clause, and will thus have to be placed on top of the stack:

46) 
(((Nyman [FIN attack]) the critics) with a)  
P1  
P2

The next word to be recognised, "a", can be linked as a dependent of "with":

47) 
(((Nyman [FIN attack]) the critics) (with a))  
P1  
P2

However, there is still not enough information here to determine what "with" modifies. Therefore no syntactic relation can be established between "with" and anything in P1, and thus the new cell in P2 will have to be placed back on top of the stack:

48) 
(((Nyman [FIN attack]) the critics) (with a) mallet)  
P1  
P2

Form-D will attach the next word in P2, "mallet", as a dependent of "a".
Now that the entire adjunct phrase has been processed, we finally have enough information to
determine that “with” must modify (assign a Subject role to) “attack” rather than “the critics”.
Thus “with” will have to be licensed passively by FIN:

Now “with” is d-commanded by “attack” and can thus assign its Subject role to it, as
illustrated in (42).

Here, then, we can see how the whole of an adjunct phrase, or at least a significant part
of it, will have to be processed before it can be linked to the rest of the clause. Needless to say,
this question will not arise with single-word adjuncts like “yesterday” and “quickly”, whose
semantic relation to a clause can be determined immediately. The important point to note here,
though, is that the interpretation of an adjunct must be worked out before the parser can
establish any syntactic relation between it and its head in the matrix clause. This implies that
the LGP, in common with many other parsing systems, will sometimes have to respond to
pragmatic as well as syntactic information (Rayner et al. 1983, Mitchell and Holmes 1985, Tait
1985).

5.3.4 - Some advantages of pair-based parsing

As should be clear from the previous discussion, the basic heuristic of the LGP is to
select pairs of adjacent words (or cells) and to establish a dependency relation between them
on the basis of their licensing requirements. In this way the parser will always seek to build
syntactic structure by examining the properties of two words at a time, hence the two-place
list which defines the application of Form-D. As I pointed out in section 5.3.1 this is unusual,
and according to Fraser (1993) the only other dependency-based parser to adopt a similar pair-
based strategy is Hays’ (1961) somewhat rudimentary parsing system. As far as I am aware,
all other dependency parsers employ a search-based strategy which, put very simply, tends to
involve accessing the dependency requirements of one word at a time and then seeking other
words in the input string which may satisfy these requirements.
In many cases the differences between a pair-based parser like LGP and a search based parser are rather subtle, and there are undoubtedly parallels between the two systems. For example in some cases a search-based parser will have to ‘look for’ words which have already been processed and added to the list, and this is essentially similar to the operation of a pair-based system. Whatever the similarities between the two, however, I believe that there are certain circumstances in which a pair-based parser can be shown to enjoy specific advantages over a purely search-driven strategy. One such case involves structures such as the following:

51)  
i. Mike was eating.  
ii. Mike was eating a fig.

Verbs like “eat” display a systematic valency ambiguity between a transitive and intransitive reading which could cause problems for a purely search-based parser; on encountering the verb “eat” a search-based processor will be unsure whether or not to search for a dependent. In effect, we only know whether “eat” is transitive or intransitive once an object is actually encountered or confirmed missing. Here, though, by admitting that the parsing process may be driven by the mere occurrence of another word, we have, in effect, diluted its search-driven orientation. Examples like (51) thus point to something more similar to the pair-based heuristic of the LGP in which processing is driven by the requirement to establish syntactic relations on the basis of two words rather than a single word.

Perhaps the most important advantage of the pair-based LGP heuristic is that it can be shown to derive the ‘syntactic’ principle of adjacency. Consider the structure in (52):

52)  
\[ \text{*Ted [FIN crash] his yesterday Bentley} \]  
\*‘Ted crashed his yesterday Bentley’.

As I pointed out in chapter 3, structures of this sort are generally ruled out in dependency theories by an independent syntactic principle of adjacency which prevents a dependent from being separated from its head by other words which bear no relation to them (Robinson 1970, Hudson 1984-a, 1990, 1994, 1996, Mel’čuk 1988). In this way the adjacency principle can be viewed as a well-formedness constraint, the overall effect of which is to exclude structures with crossing or ‘tangling’ dependencies (see section 3.2.3 of the third chapter).

---

7A similar point can also be made in relation to post-adjuncts.
However, I believe that the adjacency (or ‘no tangling’) principle need not be stated as a separate grammatical rule of LG, but can instead be shown to follow naturally from the pair-based operation of the LGP. (53) below is a schematised version of the ungrammatical structure in (52):

53)

*A B C D

From this structure we can infer that A and B will each license one dependent, as well as needing a head. In parsing (53), the first two words to be recognised will be A and B; these are placed on P1 and P2 respectively:

54)

A

_____B

PI P2

As before, there are two possible outcomes of Form-D’s application here; given that A licenses a dependent and B needs a head, if B conforms to any categorial constraints imposed by A then Form-D will operate successfully and establish a dependency between the two:

55)

(A B) C

PI P2

Note however that this option is already incompatible with the structure in (53) since A’s licensing capacity has been expended on B and will consequently be unavailable for C. Alternatively, referring back to (54), if B does not conform to A’s categorial requirements, Form-D will fail to operate, and B will be placed on top of A in the stack in P1, making way for the next element, C, to be placed in P2:

56)

A / B \\ C

PI P2

Given that A is beneath B in the stack and thus invisible to the processor, the only possible application of Form-D will be to link B, at the top of the stack, with C. Once again there are two possible outcomes; either Form-D will link B and C, creating a cell in P2 (57i), or Form-D will fail, and C will be placed on top of B in the stack in P1 (57ii):
Once again, however, both of these options are incompatible with the structure in (53); there is simply no way in which A's licensing capacity could bypass B and remain open for C. Similarly, the pair-based operation of the parser guarantees that any attempt to allow B to license D 'across' C will also fail. I therefore believe that the sentence in (52) may after all be 'grammatical', in the sense that it is generated by the syntax (the lexicon and the head parameter), although it can be ruled out on the grounds that it is difficult to process. Of course the sentence in (52) is not totally unprocessable. Nevertheless, any strategy required to process this structure will run contrary to the general pair-based parsing procedure of the LGP.

This last point is of particular importance since it represents another fundamental departure from much of what is conventionally assumed in other dependency grammars and their related parsing systems. As I pointed out above, some form of adjacency constraint is generally stated as an independent syntactic principle in dependency grammars, and this principle is then allowed to constrain the operation of the parser. For example Hudson (1989, 1994) and Fraser (1993) show how a syntactic adjacency constraint can control the operation of a search-based parser by preventing it from looking for a dependent which is not next to its head. What I am suggesting is the opposite of this; instead of using a syntactic principle of adjacency to control the parser, I believe that a version of this principle can be shown to fall out as a natural consequence of the LGP's operation. If, then, adjacency does not have to be stated as a separate principle of LG syntax, it represents one further way in which LG can be shown to be 'smaller' than other syntactic theories.

5.3.5 - Cell structure and reanalysis

As I have already noted, another of the slightly unusual aspects of the LGP is the way it builds up cell structure incrementally as words are processed. This cell structure serves to record the order in which dependencies are established by the parser. Consider once again the structure in (33ii) above, repeated here with labelled cells:

58) \[ (c_{_{_{\text{the critics}}}_A}) \rightarrow (\text{FIN} \text{ despi}e_{_{_{B}}}) \rightarrow \text{Nyman}_C) \]

'The critics despise Nyman'.
The first relation to be established by the parser, between “the” and “critics”, is registered in the innermost cell, marked A here. The next link to be made, between “the” and the FIN element, is encoded in the ‘middle’ cell marked B, while the final relation that Form-D established, between “despise” and “Nyman”, is registered in the outermost cell, marked C. The same system can also be applied to the more complex structures in (40) and (50) above. In this way the cell structure built up by the LGP during processing is reminiscent of the concentric rings within a tree trunk, with each successively more embedded cell (or ring) representing an earlier stage of processing (or growth). In this way the entire processing history of a sentence can be reflected in its cell structure.

The question is whether this permanent record of a structure’s processing history is relevant. I believe that it is, for a number of reasons; in the next section, for example, I will show how cell structure plays an important part in accounting for extraction. I also believe that the derived constituency implied by cell structure may be valuable in accounting for co-ordination phenomena, although I do not have the time to discuss this issue further. Here, though, I want to concentrate briefly on another use of incremental cell structure, in constraining the application of backtracking and reanalysis (Frazier and Rayner 1982, Marcus et al. 1983).

As I noted in section 5.3.1 the LGP is a serial parser which seeks to create a single structure when faced with an ambiguous input. Since, however, the LGP does not make use of any look-ahead facility (Marcus 1980), there will inevitably be occasions where ambiguous structures will cause the parser to create an incorrect structure:

59) We know Santa stole the cash.

As I have already pointed out, sentences like (59) display a temporary local ambiguity in that when processing the third word, “Santa”, an incremental parser lacking any relevant controlling heuristic will be unsure whether to attach it as an object of the matrix verb “know” or a subject of an embedded clause. Consider now how the sentence in (59) would be

---

^Bearing in mind the diachronic significance of cell structure, in some sense the word “despise” in (58) could be described as the innermost cell of all; The two elements of this word, FIN and the verb, are linked by a licensing relation which evidently ‘predates’ any parsing process. In this way the square brackets of a two-element word could be taken to represent a ‘fossilized’ cell, of which the licensing relation is not a product of parsing but of lexical fusion (see section 4.3.5 of chapter 4).

^Both Tesnière (1959) and Hudson (1988-a, 1990) concede that some form of constituency may be the only plausible way of accounting for co-ordination in a dependency grammar.
processed by the LGP.

The first two words to be recognised will be processed as usual, with “we” being linked to the FIN element of “know”. The next word to be recognised and placed in P2 is “Santa”:

60) \[
\begin{array}{c}
\text{we} \quad \text{[FIN know]} \quad \text{Santa} \\
P1 \\
\text{we} \quad \text{[FIN know]} \quad \text{Santa} \\
P2
\end{array}
\]

Since “Santa” is a nominal which requires a head and “know” is a verb which can license a nominal dependent, in the absence of any evidence to the contrary, Form-D will apply here in the only way it can, and respond to the dependency requirements of these two words:

61) \[
\begin{array}{c}
\text{we} \quad \text{[FIN know]a} \quad \text{Santa} \quad \text{[FIN steal]} \\
P1 \\
\text{we} \quad \text{[FIN know]a} \quad \text{Santa} \quad \text{[FIN steal]} \\
P2
\end{array}
\]

However, incorporating “Santa” into the cell structure in P1 has made way for the next word “stole” in P2. At this point it is clear that the parser has made a mistake; there is nothing available in P1 to license “stole” actively, and clearly this finite verb would be unsuitable as a passively-licensed adjunct.

In order to salvage the situation in (61) the parser will be forced to backtrack and undo some of the structure it has created. There are, of course, two relations which could be ‘unpicked’ in (61), either the dependency between FIN and “we” or that between “know” and “Santa”. In practise, though, we know that it will be the latter which is undone so that “stole” can be linked as a dependent of “know”. I believe that the incremental cell structure in (61) may help explain why this is the case. Simply, it seems reasonable to assume that the first relation to be undone through backtracking will be the last one to have been established by the parser. This, of course, is precisely what the cell structure encodes; the fact that the last dependency to be established by Form-D was between “know” and “Santa” is recorded in the outermost cell, marked B in (61). It is thus appropriate that the most vulnerable dependent for reanalysis should be situated in the most external, ‘exposed’ cell. The parser will thus target this outermost cell for backtracking and reanalyse “Santa” as a dependent of the FIN element of “stole”. This same finite verb can then itself become a dependent of the matrix verb “know”.

However, it is possible to make a more interesting claim, and assume that the parser should not generally backtrack beyond the outermost cell; any processing difficulty associated with backtracking could then be predicted to increase proportionally with each cell that is
bypassed beyond the outer cell. In this way processing complexity might be computed directly according to how far ‘inside’ a structure a revaluation must be carried out. This of course reflects the natural intuition that more established relations formed earlier on should be safer from reanalysis than those which were formed later. I do not have the time here to examine the implications of this possibility more thoroughly. However, it does appear to be the case that ‘garden path’ clauses which commonly cause processing breakdown do involve compulsory reanalysis of relations encoded within cells which are fairly deeply embedded inside the structure.

Apart from constraining the operation of backtracking, cell structure also provides a novel way of representing the way in which reanalysis may take place. In (61), for example, rather than employing a destructive mechanism to obliterate the dependency relation (and cell) between “know” and “Santa”, we can simply say that the external B-cell remains as it is and that “stole” ‘moves in’, superseding the original dependent “Santa”. In this way “stole” could be said to ‘usurp’ “Santa”’s position as the dependent of the B-cell. Of course it will then be up to the FIN element of the ‘incoming’ transitive verb to license “Santa”, the newly detached word whose original head it has usurped. I assume then that in order to salvage the situation in (61) the parser will target the B-cell of the structure in P1 as the most vulnerable for reanalysis (62i), and will then move the new word, “stole”, into this cell (62ii):

![Diagram](image)

In this way the parser is able to carry out a reanalysis leaving the cell structure intact. Processing will now continue as usual until the subsequent words, “the” and “cash”, have also been attached to the structure in P1.

---

\[^{10}\text{The curly brackets enclosing the C-cell here reflect its exceptional status in having been created inside an existing cell, B. This of course goes against the usual diachronic pattern of cell structure, and I assume that reanalysis is the sole circumstance in which this can occur.}\]
5.4 - Extraction and Valency Transfer

5.4.1 - LG and the syntactic basis of extraction

Consider once again the structure in (27) repeated as (63) below:

63)

\[ \begin{array}{c}
\text{The critics} \quad \text{[FIN despise]} \quad \text{Nyman} \\
\end{array} \]

‘The critics despise Nyman’

Following the discussion in section 5.3 we know that the parser will have no problem here in determining that “the (critics)” must be the licensee of FIN and that “Nyman” must be the licensee of “despise”. On the basis of (63), then, we might assume that the structure for an analogous sentence with a ‘topicalised’ object would be as shown in (64):

64)

\[ \begin{array}{c}
\text{Nyman} \quad \text{the critics} \quad \text{[FIN despise]} \\
\end{array} \]

‘Nyman the critics despise’.

However, as it stands the structure in (64) violates the revised head parameter discussed in the previous chapter and repeated in (31) above. According to this only FIN is allowed to have a predependent, whereas in (64) the verb “like” is also shown to have a predependent. If this structure is to obey the head parameter it will thus have to be revised as follows:

65)

\[ \begin{array}{c}
\text{Nyman} \quad \text{the critics} \quad \text{[FIN despise]} \\
\end{array} \]

However, this analysis is equally problematic since now FIN is shown to actively license three dependents while only actually specifying two. More importantly, the structure in (65) fails to reflect the fact that “Nyman” is evidently a syntactic and semantic argument of the verb “despise”, and indeed the syntactic capacity of this verb to license a dependent remains unfulfilled here. We are therefore faced with a conundrum in that we must somehow allow “despise” to license “Nyman” while not falling foul of the head parameter.

It might initially appear that we could avoid this problem by allowing FIN to license “Nyman” passively. However from the discussion in section 4.4.6 of the previous chapter it should already be clear that “Nyman” is an unsuitable passive dependent since it assigns no semantic role and thus cannot guarantee its own semantic integration into the sentence. Moreover, even if FIN were to license “Nyman” passively, the licensing properties of “despise”
would still remain unfulfilled. I believe that a more attractive solution to the problem posed by (64)/(65) would be to allow the unfulfilled licensing capacity of “despise” to be transferred to FIN. In this way FIN could license “Nyman” actively through a valency specification ‘borrowed’ or transferred from the verb. In (66) below this valency transfer is represented by the ‘square’ arrow pointing from “despise” to FIN:

66)

Here then FIN actively licenses “Nyman”, in accordance with the head parameter. However this additional licensing capacity actually originates from “despise”; in other words “despise” is the source of the licensing dependency which links FIN with “Nyman”, thus reflecting the fact that “Nyman” is actually a syntactic and semantic argument of this verb. In order to make this point clearer we could identify the dependent and the source of a transferred licensing relation by means of a shared index:

67)

In this way FIN, the recipient of the transferred licensing relation, can be shown to play something of an intermediary role between the source (“despise”) and dependent (“Nyman”).

I will refer to this system of shifting words’ licensing capacity as valency transfer (VT), and in the rest of this chapter I will explore the possibility that this may constitute the central syntactic mechanism underpinning LG’s account of extraction. The operation of VT can be expressed by a simple rule, stated in (68) below, where X, Y and Z represent variables over a variety of potential categories or structures:

68) - Valency Transfer (VT)

Thus, through VT, X in (68) could be said to acquire a capacity to license two dependents (Y and Z) because it takes over Y’s ability to license its own dependent. I assume that this type of configuration can then also generalise to cases of long-distance extraction, as illustrated in (69) below:
The 'square' arrow indicating VT in the above examples should not be confused with the arrows representing licensing relations; "despise" in (67), for example, does not license FIN. However, the use of an arrow to represent VT does help reflect the fact that what is transferred is, in effect, a licensing relation.

I assume that the VT rule in (68) will be stored as part of the grammar alongside the lexicon. Indeed, apart from the head parameter, this is the only non-lexical rule so far to be included in LG. It is important to bear in mind, though, that in spite of superficial similarities, the VT rule in (68) does not represent a transformation; one structure is not transformed into another. Instead (68) merely serves to express that the two structures are equivalent. It is true that the term 'transfer' implies that something starts off in one place and ends up in another. I do not, however, intend this to express any process of derivation within the syntax. Any such notion of syntactic derivation would in any case be incompatible with the monostratal nature of LG. Nevertheless, the derivational overtones of the term 'Valency Transfer' are not entirely inappropriate since, as I will show, even though VT is a static syntactic phenomenon, its application will be executed 'online' by the parser.

5.4.2 - Constraining VT

Initially it may appear that VT, in allowing licensing specifications to be transferred from one word to another, may introduce an unacceptable degree of unconstrainedness into LG; after all, the implication of the rule in (68) is that any word in a structure might be allowed to license anything else simply by acquiring the licensing properties of another word in the same structure. The ultimate result of this would of course be that LG overgenerated massively, and it would thus be impossible to discover the syntactic structure of any sentence. The application of VT evidently has to be constrained in some way so that its application is restricted to a small class of specific and recognisable circumstances. One way of achieving this would be to modify the rule in (68) so as to ensure that it only applied in those cases where it was required. However, this is an unattractive solution since any rule that was sufficiently specific to cases of extraction would amount to little more than a statement that extraction can occur in language (Chomsky 1981, 1986-b). Moreover, even if it were possible to formulate
a single rule capturing all the intricacies of extraction data, the result would in all probability be huge, cumbersome and unlearnable. I believe that a more promising approach is to leave the rule in (68) as it stands and seek to constrain its application by independent principles. In this way VT could be left to apply freely and randomly as a syntactic operation, very much like Move-α in PP Theory described in section 2.3.1 of the second chapter. It will then be up to incidental factors to constrain the application of VT in such a way that it will only apply in exactly those cases where required.

However, unlike Move-α, I assume that VT is not constrained by independent principles of syntax, but instead by processing-related factors; although VT is a syntactic operation, it will have to be executed and governed by the parser. That is to say, it will be up to the parser to identify cases of VT and locate a suitable source and recipient word between which transfer can take place. I therefore believe that the parser is a plausible and appropriate source of constraint for VT. In fact the operation of the LGP described in section 5.3 can be shown to restrict the application of VT naturally and automatically in two separate ways, which I will refer to informally as ‘motivation’ and ‘locality’:

i - Motivation. Obviously for licensing specifications to be transferred from one word to another gives rise to more complexity than if they had remained where they should be, and it thus seems reasonable to suppose that VT will only occur when it absolutely has to. In other words, in each case of VT the recipient of the transferred valency must be forced to license something which it would not otherwise license. I assume that it is up to the parser to determine when a word is forced to license an ‘extra’ dependent, and that it will only exercise the option when the alternative would constitute an ill-formed structure. For example in (66)/(67) above FIN is ‘forced’ to license “Nyman” by the head parameter; if FIN did not license “Nyman” through VT there would be no way in which the parser could produce a well-formed structure. In (70) below, however, the use of VT is inappropriate:

70) *The critics [FIN despise$] Nyman

‘The critics despise Nyman

There is simply nothing here to motivate the transfer of the licensing specification from “despise” to FIN. Unlike the structure in (67) above, the head parameter here does not force the parser to ‘overburden’ FIN with a second dependent which it would not otherwise have. The structure in (70) is consequently unprocessable. The motivation constraint thus guarantees
that VT will only occur in the (possibly very limited) set of circumstances where a head is forced to license a dependent it would not otherwise have.

It is, however, possible to make a slightly stronger claim in relation to VT-motivation, and say that the parser can only process a VT structure once the motivation for it has been found. The implication of this claim would be that the recipient of a transferred relation, the word which is forced to license an extra dependent, must precede, or at least coincide with, the source of the transferred relation. This criterion applies in (67), for example, a case of local extraction where the source and recipient of the VT occur within the same word “despise”. The example in (69), though, illustrates how in cases of long distance extraction the recipient of VT, the matrix FIN, must precede the source (the verb “despise”). The main attraction of this stronger version of the motivation constraint is that it allows us to derive in the simplest possible way the fact that extraction only occurs from right to left (see section 1.2.1 of chapter 1). Any extraction from left to right would entail that the source of VT preceded the recipient. This would violate the stronger version of the motivation constraint since the parser would be expected to process a VT structure prior to finding any motivation for it.

ii - Locality. The rule in (68) imposes no restriction on the distance between the recipient of VT (X) and its source (Y); since X, Y and Z are variables there is no reason why they should not represent structures of infinite length and complexity, and the example in (69) illustrates the distance that may exist between the recipient of VT and its ultimate source. However, although VT as a syntactic process may be free theoretically to apply over boundless domains, the distance over which VT can actually be implemented by the LGP can be shown to be strictly limited. More specifically, there is good reason to believe that the parser can only process a VT structure if both the recipient and a potential source are visible simultaneously. In other words, at the point where the parser is forced to ‘overburden’ a word or element such as FIN with an extra dependent, there must already be a word in the same structure with an unfulfilled licensing capacity which can act as a potential source for VT to the ‘overburdened’ head. This locality constraint can be more easily expressed in terms of cell structure as follows:

---

11Extraposition and ‘Heavy NP Shift’ are sometimes cited as examples of ‘rightward movement’ (Heageman 1991). However, I do not believe that these cases are necessarily relevant to VT since it is possible to argue that they do not involve altering a constituent’s syntactic head.

12Recall that ‘source’ refers to a word with an unfulfilled licensing capacity which can be transferred to another word which is ‘overburdened’ with an extra dependent.
71) - Locality

Where C is a cell arising from linking a saturated head \((H_{\text{SAT}})\) and a dependent \((D)\) through VT, C must contain a source for VT to \(H_{\text{SAT}}\).

A 'saturated' head or \(H_{\text{SAT}}\) here refers to a head which has already licensed its dependents and can thus only license D through VT. Thus in (72i) linking D with \(H_{\text{SAT}}\) is unproblematic since the resulting cell C contains a source. In (72ii) however the processor will not allow \(H_{\text{SAT}}\) to license D since, at the time of processing, there is no potential source available:

\[
72) \quad \text{i. } (cD) (H_{\text{SAT}} X) \quad \text{ii. } *(cD) (H_{\text{SAT}} X)
\]

The locality constraint in (71) is a fairly natural one since without it the parser might be expected to initiate a VT structure without knowing whether or not any source would subsequently be found for it. By ensuring that a source is available locally prior to creating a VT structure the parser is thus able to guarantee its well-formedness. However, as I will show later in section 5.4.4, the central importance of the locality condition is that it serves to reduce single long distance applications of VT to a series of local ones. For example, in (69) at the time of linking “Nyman” with the matrix FIN element the resulting cell will clearly not contain the ultimate source of VT, “despise”, since this occurs much later in the sentence. However the locality condition in (71) forces the parser to find a series of temporary local sources, ensuring that a long-distance application of VT can only take place in a series of small steps.

The significance of the locality constraint and the way it affects the parser’s processing of VT structures will become clearer as I work through some examples in the next section. Here, though, I hope to have offered a first indication of how the syntactic operation of VT can be constrained by fairly natural parsing considerations, a desire to keep structures simple in the case of the motivation constraint, and a need to ‘see’ both participants in a VT relation in the case of the locality constraint. In this way I believe that the simple VT rule in (68) can be left free and unconstrained, rather like Move-\(\alpha\) in PP theory.

5.4.3 - Local extraction and the Priority Principle

The best way to see how the parser governs the application of VT is to work through some examples. Consider once again the structure in (67) repeated below as (73):

208
The first two words to be recognised by the parser, “Nyman” and “the”, are placed on P1 and P2 of the list respectively:

Form-D will fail to operate here, since no relation can be established between these two words; although “the” requires a nominal dependent, the only possible candidate, “Nyman”, precedes it. Any dependency between the two would thus violate the head parameter. “The” will therefore be placed above “Nyman” in the stack in P1. This leaves P2 free to accommodate the next word to be recognised, “critics”:

Here the licensing properties of “the” can be satisfied, and Form-D will thus remove it from the top of the stack and link it with “critics” creating a new cell in P2:

Once again, however, no dependency can be established between “Nyman” and the cell in P2, and this cell will therefore be placed in the stack above “Nyman”, leaving P2 free for the next word “despise”:

As before, both FIN and the verb “despise” require a dependent, though the head parameter will guarantee that only FIN can license a predependent in P1. The only part of P1 that is actually visible to the processor in (77) is the cell (the critics), which is at the top of the stack. Form-D will thus establish a relation between “the”, the head of this cell, and FIN:
At this point the parser faces a problem; it is left with a word in P1 which has not been licensed, whereas the valency of FIN, the only head capable of licensing a predependent, has already been satisfied by "the critics". The only way to overcome this problem while conforming to the head parameter is to allow FIN to license "Nyman" with a licensing capacity acquired from another word through VT. The motivation constraint on VT has thus already been satisfied, since the alternative to linking "Nyman" to FIN would be an ungrammatical structure which violated the head parameter. However, in accordance with the locality constraint stated in (71), before linking "Nyman" as a dependent of FIN, the parser must ensure that the cell arising from this relation will contain a potential source for VT to FIN. In this case the resulting cell will, of course, contain the verb "despise" which, having an unfulfilled licensing capacity of its own, will qualify as a source. The parser can thus proceed with a local application of VT:

Here, then, we have a well-formed structure involving VT. "Despise" has been identified as the source of the transferred dependency relation which links "Nyman" with FIN, thus reflecting the fact that "Nyman" is a syntactic and semantic argument of this verb. The principles of Subject and Object identification can now apply normally in deriving a semantic representation from this structure:

---

13Note that although the structure in (79) is grammatical, it required six stages of processing derivation, shown in (74)-(79). The exact equivalent non-topicalised structure in (33ii), however, had only four stages of processing, shown in (28), (30), (32i) and (32ii). The additional processing effort required for extraction structures using VT such as (79) might thus help account for the increased level of markedness associated with these constructions.
Now consider the ungrammatical structure in (81):

81) *Nyman the critics despise composers.

Obviously, this structure can be excluded for simple syntactic reasons; there are too many dependents and not enough heads to license them. Nevertheless, it is worth examining how (81) would be processed. On the basis of the first four words the parser will construct the same grammatical structure as (79) above. However, in the case of (81) there is another word to be processed in P2, “composers”:

82) 

Evidently the parser will have to try to continue operating here in seeking to link “composers” with the rest of the structure in P1. One possibility is that Form-D will fail, since “despise” has already transferred its valency to FIN and thus cannot license any ‘proper’ dependent of its own. “Composers” will thus remain unconnected to the structure in P1 since there is nothing there to license it. This would of course account for the ungrammaticality of (81). I assume, though, that the parser will not operate this way but will instead allow “despise” to license “composers”:

83) 

The result of this is that “despise” no longer qualifies as a source of VT to FIN since it now has no spare licensing capacity to transfer. FIN will thus only be able to license “Nyman” passively, like an adjunct. Once again this accounts for the ungrammaticality of (81), since “Nyman” assigns no semantic role of its own and is therefore unsuitable as an adjunct.

This last example illustrates an important point about the operation of the parser in executing VT. When Form-D links “Nyman” as an extra dependent of FIN in (79) and (82) “despise” is identified as a potential source for VT to FIN, in accordance with the locality constraint. However, a ‘potential source’ means exactly that; the verb “despise” might be a source for VT in (79)/(82) because, at the time of processing, it has no dependent of its own. If, though, at a later stage a ‘real’ dependent is discovered for “despise”, such as “composers” in (83), then the verb will evidently no longer qualify as a potential source, and the resulting
structure will be ill-formed. In the grammatical example in (79), of course, no 'real' dependent is subsequently found for "despise", and thus this verb can retain its status as a source of VT. Generally, then, a word will only be a source of VT by default, so long as it has no true dependent of its own. We may formalise this by means of the Priority Principle:

84) - The Priority Principle

A word is only a potential source for VT in the absence of any 'real' dependent of its own.

The Priority Principle is a governing heuristic of the LGP, ensuring that a word's capacity to license its own dependent will have priority over its capacity to act as a source for VT. In this way each source for VT will be inherently provisional until proven otherwise. Here, then, we see the crucial role that parsing plays in administering VT; it is only through examining how a structure is processed that we can take into account the fact that a word's licensing capacity will be temporarily unfulfilled until it is linked with its dependent. In this way any word, prior to being attached to a dependent of its own, can qualify as a provisional source of VT. As I will demonstrate in the following section, this is of particular importance in cases of long distance extraction.

5.4.4 - VT and long-distance extraction

The sentence in (85) below displays the long-distance extraction of the object “Nyman”:

85) Nyman I know that critics despise.

In processing this structure, the parser will initially place “Nyman” and “I” in P1 and P2 of the list:

86) 

Since no relation can be formed between these two words, “I” will be placed on top of “Nyman” in the stack in P1 leaving P2 free to accommodate the next word, “know”:

87) 

212
Once again Form-D will be able to operate successfully on “I”, at the top of the stack in P1, and FIN, the only element of the word in P2 which is capable of licensing a predependent:

88)  
   \[ \text{Nyman} \quad \text{(I \ [FIN \ [\text{know}])} \quad \text{FIN} \quad \text{P2} \]

The processor is now placed in the same situation as before in that it is faced with a headless dependent, “Nyman”, in P1 while the only head in P2 capable of licensing a predependent, FIN, has already expended its licensing capacity on “I”. Once again, then, the processor will be forced to let FIN license “Nyman” with a licensing specification acquired through VT. The cell arising from linking “Nyman” with FIN will of course contain “know”, a verb with an as yet unfulfilled licensing capacity which can thus act as a source of VT to FIN. Of course “know” will not be the ultimate source of VT in this structure, but at the stage in processing in (88) the incremental LGP is not in a position to know this. In accordance with the locality constraint, then, “know” will be identified as a local potential source, and the parser will therefore proceed with a VT structure:

89)  
   \[ \text{Nyman} \quad \text{(I \ [FIN \ [\text{know}])} \quad \text{that} \quad \text{P2} \]

The next word to be recognised and placed in P2 is “that”. The Priority Principle in (84) ensures that this will be linked as a dependent of “know”, the result being that this verb will no longer be a source of VT to FIN (90):

90)  
   \[ \text{((Nyman \quad (I \ [FIN \ [\text{know}])} \quad \text{that}) \quad \text{critics} \quad \text{P2} \]

It may initially appear that the structure in (90) will be ungrammatical in the same way as (83) above, since the original VT to FIN is left without a source. However, I do not believe this is true; unlike “composers” in (83) the nominal which is added to the structure here, “that”, has its own (unfulfilled) licensing capacity, and I believe that it can therefore be identified as a new source of VT to FIN. In other words, at the point where the original VT relation between “know” and FIN is broken off by the attachment of “that”, a new VT relation can immediately be established between “that” and FIN, thus rescuing the structure:
In this way the parser can create consecutive VT structures, with one legitimately replacing the other. Naturally this new VT structure will have to obey the locality constraint; in (90) the word which disrupts the original VT structure and thus requires a new one to be formed is "that". As (91) shows, though, the cell which arises from linking "that" to "know" will, evidently, contain "that" itself, which, having an unfulfilled licensing capacity, will thus qualify as a local source for VT. In the case of the ungrammatical structure in (81)/(83) above, however, no such source was available since the nominal which interrupted the original VT structure, "composers" brought with it no licensing capacity of its own. In other words (83) could be said to violate the locality constraint in that the parser is unable to find any alternative local VT-source to replace "despise".

Returning now to (91), needless to say "that" would not be suitable as the ultimate source of VT to FIN; if the structure in P1 were to remain as it is then it would be ungrammatical since "that" cannot license a nominal such as "Nyman", even through VT. Given the temporary nature of some VT structures, though, I assume that "that", by virtue of its spare licensing capacity, will be a valid transitory source for VT until a new one can be found. Certainly in the case of (91) the parsing process will have to continue, prompted by the presence of the next word "critics" in P2. Form-D will be unable to link this with anything in P1; although the Priority Principle would usually determine that "critics" was linked as a dependent of the VT source, "that", we already know that "that", as a complementiser, cannot license a nominal dependent. Thus the structure in P1 will remain as it is and "critics" will be placed on top of it in the stack:

\[ 92) \]

The next word in P2 is "despise". The FIN element of this word can be linked to "critics" at the top of the stack in P1, creating a cell in P2:

\[ 93) \]
At this point the cell in P2, headed by FIN, constitutes a suitable dependent for “that” in P1. In accordance with the Priority Principle Form-D will attach FIN as a dependent of this word, thus severing the VT relation between “that” and the matrix FIN. Once again, however, as with the structure in (91), the word which is responsible for breaking off the VT relation here, “despise”, has an unfulfilled licensing capacity of its own and will thus itself qualify as a local potential source for a new VT relation. Thus the parser can again proceed with a new consecutive VT structure in accordance with the locality constraint:

94)

\[
\text{(((Nyman, [FIN know])) that) (critics [FIN despise]))} \rightarrow \text{P2}
\]

In (94) we are left with a well-formed structure in which “despise” has finally been identified as the source of the transferred dependency relation linking the matrix FIN with “Nyman”. In this way “Nyman” can be described as a syntactic and semantic argument of “despise”. However, although “despise” is the ultimate source of VT in (94), it is the third such source to be found within the structure. Initially “know” was identified as a VT source in (89), followed by “that” in (91). These two other temporary VT relations could thus be said to mediate the long-distance VT relation between the matrix FIN and “despise”:

95)

\[
\text{(((Nyman, [FIN know])) that) (critics [FIN despise]))} \rightarrow \text{P2}
\]

It was only through these extra (temporary) VT structures that the processor was able to operate locally, thus yielding a grammatical structure. For example, at the stage in processing shown in (89) the verb “know” was the only possible local source of VT in the structure. In this way I believe that long distance applications of VT can be localised and constrained by the operation of the parser. Note, though, that this localisation does not have to be stipulated but follows naturally from the parser’s need to find a local source for each application of VT.

The structure in (95) is reminiscent of the WG analyses of extraction involving the visitor relation (see, for instance, example (46) in section 3.4.2 of the third chapter). In both cases the extractee could be said to bear a series of relations to heads and their dependents throughout the sentence. It is important to remember, though, that there are important differences between the representation in (95) and the WG analysis. For one thing (95) does
not represent a syntactic structure so much as a processing history of the structure in (94); the verb "despise" is really the sole true source of VT in these examples, and the additional applications of VT in (95) are really only of 'historical' interest, serving to indicate the different points at which the processor located temporary potential sources during its derivation. In the WG analysis, though, the corresponding extra visitor relations which link an extractee and different heads are very much part of the syntactic structure of the sentence.

Even abstracting away from the diachronic significance of the representation in (95), important differences remain between this and the WG analysis. For example, the long-distance relation in (94) which is localised in (95) is not a dependency but a transfer of licensing capacity. This has some important implications, notably in the fact that VT, unlike a dependency relation, is not subject to any form of adjacency constraint, whether stated as a syntactic principle or derived from the operation of the parser. Thus the diachronic localisation of long-distance VT relations such as in (95) is not achieved so as to avoid a violation of any adjacency constraint, but is simply an incidental reflex of the LGP's local application of VT. Another notable difference between the VT structure in (95) and the WG analysis is that the latter invokes a specific and distinct GR in the form of the visitor relation. The VT structure in (94)/(95), though, is composed entirely of the same type of licensing relation found in any other structure; the dependency between the matrix FIN and the extractee "Nyman", for instance, is essentially a normal active licensing relation, though one which happens to have been transferred from elsewhere in the sentence. This absence of distinct GRs in the LG account of extraction is, of course, to be expected given the theory's 'mono-relational' approach, described in section 4.2.3 of the previous chapter.

The above examples have dealt with the extraction of objects. However exactly the same principles will also apply in other cases of extraction. (96) below, for example, illustrates the extraction of a topicalised subject:

96) Nyman I know ignores the critics.

As with (85) above, on the basis of the first three words, the incremental LGP will derive the structure in (97) where the verb "know" has been identified as a source of VT to FIN:

\[\text{\ldots}\]

\[14\text{I will discuss the extraction of adjuncts separately in section 5.6.}\]
The next word, "ignore", being a finite verb, is a suitable dependent for "know" and thus, according to the Priority Principle, it will be attached as such:

\[
((\text{Nyman} (I [\text{FIN know}])) [\text{FIN ignore}]) \rightarrow \text{the}
\]

Once again the result of this attachment is that the original VT relation between "know" and the matrix FIN has been disrupted, and thus the parser will have to locate another local source of VT. In fact in (98) there are actually two local potential sources contained within the external cell, FIN and the verb "ignore". However, in terms of dependency links the subordinate FIN is closer to the matrix FIN, the recipient of the transferred valency, than "ignore" is. It therefore seems reasonable to assume that the lower FIN is 'more local' and will thus be identified as the new source of VT:

\[
((\text{Nyman} (I [\text{FIN know}])) [\text{FIN ignore}]) \rightarrow \text{the}
\]

Now processing will continue normally with the next word, "the", being linked as a dependent of "ignore", whose licensing capacity is still unfulfilled in (99). The final word, "critics" will then be linked as a dependent of "the":

\[
(((\text{Nyman} (I [\text{FIN know}])) [\text{FIN ignore}]) \rightarrow \text{the}) \rightarrow \text{critics}
\]

Here, then, just as with the long distance extraction of the object in (94)/(95), the parser is able to create a sequence of consecutive VT relations, allowing the extractee to be 'passed down' through a syntactic structure from heads to their dependents in a way which is reminiscent of WG's visitor analysis. An extracted 'topic' will initially be linked to the matrix FIN which, according to the head parameter, is the only element that can license a predependent. However, an extractee will always be linked to FIN after it has expended its own licensing capacity on the subject, and for this reason FIN will only ever be able to license
an extractee with a borrowed dependency acquired through VT. It will therefore be up to the processor to locate a source word with a spare licensing capacity which can be transferred to the ‘overburdened’ FIN head. This source will have to be visible to the parser at the time of initiating the VT relation, otherwise there would be no guarantee that any source would be found later. In this way it is possible to ensure that each application of VT is local, between words contained within the same structure at the time of processing. However, the Priority Principle, stated in (84), determines that a word which is initially identified as a potential source of VT may later prove not to be if a true dependent is subsequently found for it. In this case the parser will have to find a new source, once again, according to the locality constraint, in the same structure. This means that in order for a grammatical structure to be preserved, any ‘incoming’ word which interrupts a VT relation will have to bring with it an unfulfilled licensing capacity of its own by which it can then itself act as an alternative VT source. In this way the diachronic localisation of long distance dependencies illustrated in (95) and (97)-(100) is actually a reflex of the interaction between the Priority Principle and the locality constraint.

5.4.5 - The structure of WH-words

In the previous sections I showed how the syntactic operation of VT, when constrained by the operation of the LGP, could provide a mechanism underpinning the local and long-distance extraction of topicalised elements. In this section I will examine how the same principles may apply to the extraction of WH-elements. First, though, it is necessary to examine the internal structure of WH-words.

WH-words can serve either as an argument or an adjunct within a clause:

101) i. What did Santa steal?
    ii. Why did Wally wear a fez?

However, in some sense these WH-words also function as the heads of their respective clauses; recall, for example, the evidence cited in section 3.4.3 of the third chapter. In 1.3.3 of the first chapter I even suggested that a WH-word can be thought of as a type of complementiser; certainly WH-words serve to ‘introduce’ their clause in a similar way to words such as “that” and “if”. How, then, can a single word simultaneously function as both the head of a clause and an argument or adjunct of a constituent of that same clause? In section 3.4.3 of the third chapter I described how the WG approach to this question is to allow a mutual dependency to exist between the WH-word and the main verb of the clause, whereby the WH-word is both
the head and a dependent of the matrix verb. However this option is not applicable to LG for
the simple reason that structures involving mutual dependencies are inherently excluded in
terms of licensing; X cannot sanction Y's occurrence and simultaneously be sanctioned by Y;
as I noted in section 4.2.2 of the previous chapter, the concept of licensing entails a degree of
logical priority between a licenser and its licensee. Once again, then, the more constrained
nature of licensing which lies at the heart of LG's conception of dependency forces us to seek
another approach to the issue.

However there is another possible solution to this problem; in section 4.3.5 of the
fourth chapter I discussed how in LG a single word could be composed of more than one
element, and here I would like to examine the possibility that WH-words, like finite verbs and
gerundives, may also consist of two elements. One element could serve as a 'complementiser'
introducing and licensing the head of the subsequent clause, while the other element could
correspond to the argument or adjunct function of the WH-word. Let's assume, then, that each
WH-word consists of a complementiser element, which I will refer to as WH, and another
element, which I will refer to as α, which serves either as an argument or as an adjunct of a
word in the subsequent clause:

102) [WH α]

Very simply, what I am suggesting is that the basic structure of a WH-sentence might be as
shown in (103), where the WH complementiser licenses FIN, the head of the clause:

103) i.

[WH α] [FIN do] Wally wear

'What did Wally wear?'

ii.

[WH α] [FIN do] he wear it

'Why did he wear it?'

Following the assumptions outlined in section 4.3.5, if we are to assume that two
elements of the WH-words in (102) and (103) constitute a single word, then they themselves
will evidently have to participate in a licensing relation. It is reasonable to assume that the WH-

---

I assume here that "do" passively licenses α so as to allow α to be d-commanded by, and thus
assign a Subject role to, "wear".
complementiser element, being the supposed head of the entire clause, will license the α-argument/adjunct element rather than vice versa. In this way the internal structure of a WH-word will be as follows:

104) \[
\begin{array}{c}
\text{WH} \\
\alpha
\end{array}
\]

I assume that α is a variable over an unspecified argument or adjunct. If α serves as an argument then it will have to be actively licensed and assigned a semantic role. If, though, α functions as an adjunct it will be passively licensed and assign its own semantic role, in common with other adjuncts as described in section 4.4.6 of the fourth chapter. As should be clear from the representations in (103), though, this active or passive licensing of α will be performed by some other word in the subsequent clause, suggesting perhaps that α might in fact have two heads:

105) \[
\begin{array}{c}
\text{WH} \\
\alpha
\end{array} \quad \text{[FIN do]} \quad \text{Wally wear}
\]

‘What did Wally wear?’

This double-headed analysis of α is unattractive, however, for a number of reasons which I will not discuss further here. Certainly the licensing properties of the WH and α elements are not always as indicated in (105). For example, there are cases where the WH ‘complementiser’ element does not serve to introduce a clause at all. These include ‘sluicing’ constructions (106i), \textit{in situ} constructions (106ii) and one word WH-interrogatives (106iii):

106) i. Santa hid the cash, but I don’t know where
   ii. Santa hid \textit{what}?
   iii. A: Santa’s on the run.
       B: What?!

Here we must assume that the WH element will only license α and not FIN, contrary to what is implied by (105). Similarly, the α element of the WH-words in (106) will evidently only be licensed by the WH element and will have no second head.

What we can see from (105) and (106), then, is that WH-words can behave in two distinct ways; the WH element may introduce a clause and license FIN, in which case the α element will be licensed (actively or passively) by another word in this same clause.
Alternatively, the WH element will not license FIN, and will instead only license the \( \alpha \) element of the same word. In these cases the WH-word can function as a self-contained unit, as illustrated in (106). Ideally, then, we should seek to explain this variation in the behaviour of WH-words as well as the close interaction between the putative licensing properties of their constituent WH and \( \alpha \) elements.

I believe that the VT mechanism outlined in the previous sections may provide a relatively principled answer to this question. First, though, it is necessary to make a stipulation concerning the licensing properties of the WH element. Very simply, I assume that the WH element may either license FIN or \( \alpha \), but not both, and that it specifies a preference to license the former rather than the latter. In other words, the WH element, being a complementiser, will basically ‘prefer’ to license FIN, but will, in the absence of FIN, license \( \alpha \). (107) below shows how the lexical entry of a WH-word might represent this preferential categorial requirement of dependent.

107)

\[
\text{WH} \rightarrow X^{\text{FIN}(\alpha)}
\]

The brackets around \( \alpha \) in (107) serve to indicate that this category is acceptable as a dependent only in the absence of a FIN element. This rather peculiar categorial specification will, of course, have to be stipulated as a specific property of WH-words. However, the fact that this stipulation is restricted to a specific class of words, and, furthermore, can be stored as a lexical property of these words, makes it relatively innocuous.

From the generalised lexical entry in (107) we can deduce that where there is no finite clause available WH will licence \( \alpha \) by default, as we would predict given that the two together constitute a single word. Assuming, then, that the \( \alpha \) element of WH-words will only require a single head, in common with all other words and elements, this will allow a WH-word to function as a self-contained unit where \( \alpha \) requires no additional licenser:

108)

\[
\text{Santa \ [FIN hide] \ [WH \ \alpha]}
\]

‘Santa hid what?!’

However, if the WH element is in a position to license FIN, then according to the lexical entry in (107), it will do so at the expense of \( \alpha \):
'What did Wally wear?'

In this way the complementiser function of the WH element could be said to be 'activated' by the occurrence of FIN after it. However, there is clearly a problem with (109); according to (107), the WH element only licenses one dependent. Since WH licenses FIN in (109), it is clearly no longer in a position to license α. As I have already pointed out, though, a licensing relation will have to be maintained between WH and α so as to preserve the integrity of the single word which they together constitute.

I believe that VT can provide a solution to this problem. Simply, I assume that after licensing FIN, WH will also be forced to continue licensing α with a valency specification acquired through VT. In this way the structure in (109) would in fact be as follows:

Thus "wear", with no 'real' dependent of its own, acts as a source of VT to WH, which allows this element to continue licensing α as well as FIN. The advantage of this analysis, of course, is that it reflects the fact that the WH-word, or at least the α element of it, is a syntactic and semantic argument of the verb "wear", while at the same time the WH element functions as a complementiser heading the entire clause. Furthermore, we can already see how the VT structure in (110) conforms to the motivation constraint discussed in section 5.4.2; if the parser did not maintain a link between WH and α through VT in (110) then the resulting structure would be ungrammatical since two elements of the same WH-word would not be linked by any dependency relation. In the next section I will describe how the VT relation in (110) obeys the locality constraint too.

The analysis in (110) has another advantage in that the licensing requirements of the WH and α elements can now be shown to remain constant, irrespective of whether the WH-word functions as a complementiser, as in (110), or as a simple argument as in (108). In both cases the WH element licenses a single dependent, either FIN or α, and the α-element has a single head, either WH or WH through VT.
5.4.6 - WH-extraction

Consider now how the structure in (110) would be processed. The first two words to be recognised, "what" and "did" will be placed in P1 and P2 of the list:

\[
\begin{array}{c}
\text{WH} \alpha \\
\text{FIN do}
\end{array}
\]

If the assumptions outlined above are correct, as soon as WH has the opportunity to license FIN it will do so; in this way the presence of FIN in P2 in (111) is enough to ‘activate’ WH’s complementiser function. Thus WH will license FIN, superseding the original dependency between WH and \(\alpha\). In order to maintain the integrity of the WH-word, the WH element will of course have to continue licensing \(\alpha\), though with a dependency relation acquired through VT. As soon as FIN is linked as the dependent of WH, the processor will therefore have to seek a local source of VT in the resulting cell:

\[
\begin{array}{c}
\text{WH} \alpha \\
\text{FIN do} \\
\text{Wally}
\end{array}
\]

In fact the cell created here contains two potential sources, FIN and the verb "do", both of whose licensing requirements are unfulfilled. Once again, though, as I described in section 5.4.4, FIN will here qualify as the most local source since in terms of dependency links it is closer to WH, the recipient of VT, than "do" is. Indeed, given that FIN is the element which engendered the need for a VT structure in the first place, it is entirely appropriate that it should be identified as a potential source for this VT. In this way the processor will transfer FIN’s unfulfilled licensing capacity to WH:

\[
\begin{array}{c}
\text{WH} \alpha \\
\text{FIN do} \\
\text{Wally}
\end{array}
\]

The next word to be encountered is “Wally”, a nominal. The only possible licenser in P1 for this is FIN, since the auxiliary “do” can only license a verb. The Priority Principle will thus ensure that FIN licenses “Wally”, which means that a new source for VT will have to be found locally. Although the new dependent “Wally” has no licensing capacity of its own, another local source is already available within the cell in (113) in the form of “do”. Thus the parser can create a new local and consecutive VT structure without any problem:
Naturally, the auxiliary verb “do” would not be suitable as the ultimate source of the relation linking WH and α, given its inability to license a nominal dependent. Once again, however, I assume that the unfulfilled valency capacity of “do” is enough in (114) to ensure that it will qualify as a temporary source of VT until another is found.

The next word to be encountered and placed in P2 is “wear”. Being a verb this is evidently a suitable dependent for “do”, and, thanks to the Priority Principle stated in (84), will be attached as such, even though this has the effect of disrupting the VT relation between “do” and WH. However, “wear” has an unfulfilled licensing capacity of its own, and will thus itself be a local potential source of VT to WH:

Processing is now complete, yielding the structure initially shown in (110) above. “Wear” here has been identified as the ultimate source of the VT to WH, capturing the fact that the WH-word “what” is (at least in part) an argument of this verb. Notice, though, that just as with the topic extraction in (94)/(95) of section 5.4.4, the ultimate VT-source in (115) was located via a series of more local, temporary sources:

The ‘diachronic’ structure in (116) thus reflects the fact that the parser was able to establish a series of consecutive VT relations with each one replacing the other directly. Once again this property stems from the fact that the parser has to execute each application of VT locally between words which exist within the same structure at the time of processing.

Naturally, exactly the same will also apply in cases of long distance WH-extraction such as illustrated in (117):

117) What did Willy say that Wally wore?
I hope that by now the operation of VT and its execution by the parser should be sufficiently clear so as to be comprehensible from the lines of processing derivation alone:

Once again then here the parser is able to replace each VT relation successively with another local one. In other words, just as with the long distance extraction of the topicalised elements described in section 5.4.4, at each stage of the derivation here there exists a potential source of VT to the 'overburdened' head, WH. I assume that this is in part what underpins the well-formedness of the structure, and in section 5.5 I will explore the possibility that an inability on the parser's part to create consecutive local VT structures in this way may lead to processing difficulty and apparent ungrammaticality.
Here, though, I hope to have illustrated how the theory of LG may account for extraction phenomena. I believe that this account is more economical than that offered by WG since it does not require any ancillary visitor-type relation, nor additional propositions governing the behaviour of this extra relation. An extraction structure in LG will consist solely of the same licensing relations encountered in any other construction, one of which, though, has been transplanted by the syntactic operation of VT. The rule governing VT, which allows licensing capacity to be transferred from one word to another, can be stated simply and generally, as in (68). However the actual application of VT can be shown to be severely constrained independently by the operation of the LGP described in section 5.3. For example, the motivation constraint, arising from the parser’s desire to keep structures maximally simple, guarantees that the parser will only ever proceed with a VT structure if it provides a grammatical alternative to an ungrammatical structure. The result of this is that in English at least VT is plausibly confined to just two circumstances, to avoid violations of the head parameter in the case of ‘topic’ extraction, and to maintain the internal structural integrity of a two-element WH-word in the case of WH-extraction.

In addition, the locality constraint, which stems from the parser’s need to ensure the well-formedness of a VT relation, also restricts the application of VT to structures which allow a local relation to be maintained constantly between the recipient and source of VT. I will say more about this in the next section. The central significance of the locality constraint, though, lies in its interaction with the Priority Principle, which states that the attachment of a word’s dependent takes priority over that word’s status as a source for VT. Together the locality constraint and the Priority Principle are responsible for ‘shuffling’ the source of a VT relation down through a syntactic structure as it is processed from head to dependent, in a way reminiscent of the WG analysis of extraction. In this way long distance applications of VT can be reduced to a sequence of local (temporary) VT structures. As I noted in section 5.2.1 and will go on to explore further in the next section, this diachronic localisation of the long-distance dependencies arising from extraction is essential in accounting for constraints on displacement such as island data.

It is true that the syntactic simplicity of the LG account of extraction is achieved at the expense of the increased involvement of the parser. However all syntactic theories will require a separate parsing system in any case, and I do not believe that any aspect of the LGP described in this chapter is any more complex or involved than any other theory of parsing.
then the operation of the LGP can be shown to derive certain important properties of extraction then so much the better. There is, I believe, no harm in allowing processing to complement syntax in this way, indeed, given the somewhat Spartan nature of LG syntax described in the previous chapter, I suspect that this sort of collaboration between syntax and processing may represent the only realistic way forward in providing a principled, holistic and flexible account for extraction within the theory.

5.5 - Constraints on WH-displacement

5.5.1 - The nature of island constraints

Given the parser’s pivotal role in administering the syntactic operation of VT, it follows that constraints on extraction could be formulated as restrictions on the parser’s ability to process VT structures. In other words, ‘island’ violations could be explained in terms of processing difficulty rather than ungrammaticality (Deane 1991, Pritchett 1991). Indeed, if, as I suggested in sections 5.4.1 and 5.4.2, VT should be left free and unconstrained as a syntactic operation, then the parsing process will inevitably be the only obvious locus of constraint on its application. In this section, then, I will argue that the ill-formedness of island violations is in fact a symptom of the parser’s difficulty in creating (local) VT structures during the processing of certain constructions.

One of the advantages of couching extraction constraints in terms of processing difficulty is that ‘difficulty’ is a relative rather than absolute term; some things can be more difficult than others. As I noted in section 5.2.1, I believe that this is a more promising approach to the notably scalar nature of island data than the rigid and absolute distinction between grammaticality and ungrammaticality implied by a purely syntactic account. Moreover, I believe that processing difficulty, unlike ungrammaticality, can be partially alleviated by pragmatic influences, allowing us a possible interface with the pragmatic criteria which sometimes affect extraction data.

My suggestion will be that the processing complexity of an island violation can be computed directly from the number of stages (or lines) in the parsing derivation at which the structure is ungrammatical. The more stages at which a structure is ill-formed during its processing, the harder it will be to parse, and thus the more ‘ungrammatical’ it will appear to be. As I have already noted, in each of the well-formed extraction structures discussed so far the parser has been able either to create a single local VT relation, in cases of local extraction,
or a series of consecutive local VT relations, in cases of long distance extraction. The result of this is that at each stage in the processing of these examples the VT structure has been well-formed in that there has been a local source for it. This is of course contingent on the fact that each new dependent that is attached to a structure, through the Priority Principle, will bring with it its own unfulfilled licensing capacity. So, for example, referring back to the case of long distance WH-extraction in (117) above, at no point in the processing derivation of this sentence, shown in (118) is the structure built up in P1 left without a local source for VT. In this section, however, I will show that the parser may not always be able to operate in this way, and that at some points in the processing of certain VT structures there may be no local source available. In these cases, even where the final structure is grammatical, the ungrammatical stages of its processing history may contribute to a sentence’s ill-formedness.

5.5.2 - The adjunct island and the co-ordinate structure island

Consider first of all the non-tensed adjunct island violation in (119ii) together with its associated structure in (119iii):

119) i. Ted drank meths [before crashing his Bentley].
   ii. *'What did Ted drink meths [before crashing _ ]
   iii. [WH α] [FIN do] Ted drink meths before [ING crash]

As I pointed out in earlier chapters, examples like (119ii) are generally felt to be ungrammatical. However, I believe that the ill-formedness of this example can be shown to arise from the parser’s difficulty in executing a local VT structure during processing. So how, then, might (119ii/iii) be processed? Starting from the beginning, the first two words to be recognised are “what” and “did”; these will be placed on P1 and P2 of the list respectively:

120) [WH α] [FIN do]
    P1    P2

As described in section 5.4.5, the presence of FIN in P2 will activate the complementiser function of the WH element in “what”. WH will thus license FIN, even though this will have the effect of undermining the original dependency between WH and α. However, since these elements constitute a single word a licensing relation has to be preserved between them. For
this reason the parser will be forced to initiate a VT structure so as to allow WH to continue licensing α as an ‘extra’ dependent. Fortunately a local potential source for VT is already available in the form of FIN itself, whose own licensing capacity is as yet unfulfilled. Thus Form-D will identify FIN as the closest potential source for the transferred licensing relation between WH and α:

121) 

The next word in P2 is “Ted”. The only head in P1 capable of licensing this nominal is FIN, since the auxiliary “do” can only license a verb. The Priority Principle will thus ensure that FIN licenses “Ted”, which means that a new local source of VT to WH will have to be found. Fortunately the verb “do” is already available within the same cell, and its unfulfilled licensing capacity will therefore be transferred to WH:

122) 

Although “do” can only license a verb, it will still qualify here as a potential source for VT to WH, albeit a temporary one. The next word to be recognised and placed in P2 is the verb “drink”. This will have to be attached as a dependent of “do”, and will then itself constitute a new source of VT:

123) 

The Priority Principle guarantees that the next word in P2, “meths”, will be attached as a dependent of “drink”, thus ‘uprooting’ yet again the VT to WH. This time there is a problem, however. When “meths” is linked to “drink” there is no new local potential source to replace the disrupted VT structure:

124) 

The structure here is ungrammatical in exactly the same way as (81)/(83) above; quite simply
there is a dependent, $\alpha$, with no head to license it properly. For this reason I have prefixed the line of derivation in (124) with an exclamation mark. Since WH receives no extra licensing capacity through VT, it will only be able to license $\alpha$ passively. This is insufficient, though, and from the discussion in section 5.4.5 it should be clear that $\alpha$, being an argument, will have to be licensed actively.

At (124), then, we could say that the parser experiences a hiatus in that it is faced with an ungrammatical structure without any obvious means of rescuing it. Even backtracking here would fail to provide any solution, since undoing the last relation to be formed, between "drink" and "meths", would then leave the latter without a head. However, the parsing process will have to continue, prompted by the presence of the next word, "before", in P2. Since there is no word in P1 requiring a dependent, the parser will recognise "before" as an adjunct which will have to be passively licensed. In accordance with the assumptions outlined in section 5.3.3, at this point the processor will lack sufficient information to attach "before" to a head, thus it will be placed on the stack above the cell already occupying P1:

125) $\text{(! } ((([WH \alpha] [FIN do] Ted) drink) meths) \text{ before [ING crash]} \text{ P1 P2 )}$

Even though "before" has a spare licensing capacity, it cannot rescue the ungrammatical structure at the bottom of the stack in P1 for the simple reason that it does not share a cell with WH, and thus does not qualify as a local source of VT. The next word in P2 is "crashing", which can be attached as a dependent of "before", at the top of the stack in P1:

126) $\text{(! } ((([WH \alpha] [FIN do] Ted) drink) meths) \text{ (before [ING crash]) P1 P2 )}$

Note, though, that the structure in P1 is still ungrammatical for the third successive line. However, now that a dependent has been found for "before", the processor can determine that this adjunct will modify (assign a Subject role to) "drink", and will thus have to be passively licensed by "do":

127) $\text{((((WH \alpha] [FIN do] Ted) drink) meths) (before [ING crash])) P1 P2 )$
Here, then, I assume that the structure is no longer ungrammatical since a word with a spare licensing capacity, “crash” is at last brought into the same cell as the overburdened WH head. In this way the parser can finally establish a new VT structure between a local source and recipient:

128) (((([WH _α]) [FIN do] Ted) drink) meths) (before [ING crash]))

The structure in (128) is of course ‘grammatical’ in the sense that the valency properties of all words are satisfied. However, there were three different stages in the processing derivation at which this structure was (temporarily) ungrammatical, shown in (124), (125) and (126). In other words there was a three-stage hiatus in processing which, I assume, accounts for the ill-formedness of the sentence in (119ii). Here, then, we can begin to see how the ‘ungrammaticality’ of an island violation might in fact equate to processing difficulty, arising from the parser’s inability to establish a sequence of local VT relations during processing.

(129ii) below offers an example of a tensed version of the same adjunct island:

129) i. Ted drank meths [before he crashed his Bentley].
   ii. *What did Ted drink meths [before he crashed _]?

I have already noted in previous chapters how violations of tensed or finite islands such as these are routinely judged to be worse than violations of corresponding non-finite islands. Given, then, the more severe degree of ‘ungrammaticality’ of the example in (129ii) over its non-tensed counterpart in (119ii) we should predict that it should be harder to process, possibly associated with more problematic lines of derivation. I assume that on the basis of the first four words in (129ii) the processor will produce the same well-formed structure as shown in (123) above where “drink” has been identified as a source of VT:

130) (((([WH _α]) [FIN do] Ted) drink) meths)

As before, the next word “meths” in P2 will be linked to “drink”, in accordance with the Priority Principle:
Once again the result of this is that there is no source of VT to FIN, meaning that the latter is unable to license \( \alpha \) actively. Just as with (124), then, the structure here is ill-formed, hence the prefixed exclamation mark. The next word which Form-D will seek to link to the structure in P1 is "before". This will be identified as a role-assigning adjunct which will be passively licensed by something in P1. Again the parser lacks sufficient information at this point to attach "before" to any particular syntactic head in the main clause. "Before" will thus be placed temporarily on the stack until it can be linked to its own dependent:

\[ 132) \]

The next word placed in P2 is "he". There is no way that Form-D can link this to "before", the only word visible in P1\(^{16} \). Thus "he" will also be placed in the stack on top of "before".

\[ 133) \]

Here, then, the structure at the very bottom of the stack remains ungrammatical for the third successive line, since a local source for VT to WH has still not been found.

The FIN element of the next word in P2, "crash" can license "he", at the top of the stack in P1, and thus Form-D will create a cell between these words in P2:

\[ 134) \]

Now the cell headed by FIN in P2 can be attached as a clausal dependent of "before" in P1. Still, though, the ungrammatical structure at the bottom of the stack remains unchanged:

---

\(^{16}\) I assume that the nominative case of the pronoun "he" will prevent the parser from linking it as a dependent of "before" here. A nominal which did not register its case overtly might of course be linked as a dependent, although this relation would subsequently have to be undone by the backtracking procedure described in section 5.3.5. Whatever the case, though, this will have no effect on the continued ungrammaticality of the structure at the bottom of the stack in P1.
Now that "before" has been attached to its own dependent, I assume that the parser will have enough information to link the entire adjunct phrase in P2 to the structure in P1; since the adjunct will modify the verb "drink", it will have to be passively licensed by "do":

Here, then, the structure is no longer ungrammatical since the verb "crash", with an unfulfilled licensing capacity, occurs within the same cell as WH. We therefore at last have a local source of VT for WH:

Once again in (137) we end up with a structure which is actually grammatical, though one which was associated with a hiatus in processing. Notice, however, that in this case the temporary ungrammaticality of the structure lasted for five lines, (131) - (135), compared to the three line hiatus associated with the violation of the non-tensed adjunct island described above. Because the tensed island violation in (129ii) is associated with more problematic lines in its processing derivation than the corresponding non-tensed example in (119ii), it is predicted to be more 'ungrammatical'. In this way it is possible to account for the ill-formedness of both examples while also explaining the contrast between them.

Exactly the same phenomenon of processing hiatus can also be shown to account for violations of the co-ordinate structure island. (138ii) below offers a non-tensed example:

138) i. Wally hates syntax and loves football.
    ii. *What does Wally hate syntax and love _ ?

The latter stages of the processing history of (138ii) are illustrated below:

17In section 1.3.1 of chapter 11 noted that violations of tensed co-ordinate structure islands could be ruled out on independent grounds, and I will therefore not discuss these structures further here.
Note that this structure, just like the non-tensed adjunct island violation in (119ii), is associated with a three-line processing hiatus, reflecting the fact that both examples display a comparable degree of ill-formedness. This parallel between the two examples rests on the assumption that “and” is a form of adjunct which cannot be attached to a syntactic head within the matrix clause until it has been linked to its dependent (see section 5.3.3).

5.5.3 - The WH-island

Violations of the WH-island can be explained in a very similar way to the adjunct and co-ordinate structure islands since these too can be shown to involve a processing ‘hiatus’. However, just as with other islands, there is also a marked contrast between tensed and non-tensed examples:

140) i. *What did he ask [where to buy _ ]?
   
   ii. *What did he ask [where they bought _ ]?

Indeed some people find non-tensed examples like (140i) almost acceptable (Chomsky 1986-a). Evidently, then, if we are to account for non-tensed WH-island violations such as this in terms of the LGP’s operation, then the prediction should be that they entail only a fairly small processing difficulty.
I assume that the parser can operate on the first four words of (140i) without any problem, yielding the grammatical structure in (141):

Here "ask" has been identified as a (temporary) source of VT to WH. The next word to be recognised, "where", will be linked as the dependent of "ask", in accordance with the Priority Principle. This disrupts the VT relation between "ask" and the matrix WH:

The absence of an alternative local VT-source to replace "ask" means that this structure is ungrammatical. I assume that the next word, "to", will be recognised as a non-finite marker, or FIN⁹. Furthermore, being an inherent complementiser, the WH element of "where" in P1 will license "to" in preference to α. Therefore the processor will be forced to initiate a second VT structure so as to preserve a licensing relation between the two elements of the embedded WH-word. However, the α element of "where" functions as an adjunct, and thus in common with other adjuncts it will only have to be licensed passively. In other words, all that needs to be transferred to the embedded WH element so that it can continue to license α is a passive, and thus somewhat nebulous, licensing capacity¹⁸. Thus any word, irrespective of whether or not it has any unfulfilled active valency, will trivially qualify as the source of a 'passive VT', and a local source for this new VT to the embedded WH can therefore be found immediately in "to" itself:

It will of course be up to the semantics to determine which head will ultimately act as a source of the transferred passive licensing relation between the lower WH and α; being an adjunct,

¹⁸For the moment it will, of course, have to be stipulated that any passive relation originating from WH itself, rather than being transferred to WH through VT, will not suffice to license the α element here. However in section 5.6.1 I will offer an explanation for this.
the $\alpha$ element of "where" will assign a Subject role to whatever it 'modifies', and thus its VT source allow it to be d-commanded by its Subject$^{19}$.

For now, though, the important point is that a new VT structure has been established with "to" as its source. Given, though, that "to" has only transferred a passive 'valency' to the lower WH, the same word will evidently also be in a position to transfer its own unfulfilled active licensing capacity to the matrix WH. In other words, "to" can be identified as a local source of VT to the matrix WH, thus rescuing the ungrammatical structure in (143):

At this point "to" is the source of two VT relations, one active to the matrix WH and the other passive to the embedded WH. However, the Priority Principle determines that the next word in P2, “buy”, will be attached as a dependent of “to”, thus cancelling its status as a source of active VT to the matrix VT. This is not a problem, however, since the new dependent “buy”, with its own spare licensing capacity, is itself able to undertake the role of source for this VT:

Here then we have arrived at a grammatical structure which involves two separate VT relations. Notice that this structure was associated with two ungrammatical lines of derivation, shown in (142) and (143), one less than the three associated with the non-tensed adjunct violation discussed in the previous section. This, I think, adequately reflects the fact that violations of a non-tensed WH-island are generally accepted to be less severe than violations of a corresponding non-tensed adjunct island; compare, for example, (140i) with (119ii).

Now consider the tensed version of the WH-island in (140ii) above. As I have already noted, this is clearly considerably worse than the corresponding violation of the non-tensed structure in (140i), and we should thus assume that it will be associated with additional ungrammatical lines in its derivation. I assume, once again, that the first four words of (140ii)

---

$^{19}$In the case of (143), it so happens that "do" will ultimately turn out to be the correct source of VT to the lower WH since this allows the $\alpha$ adjunct element of "where" to be d-commanded by, and thus assign a Subject role to, the verb "buy". 

236
can be processed without any problem yielding the grammatical structure in (146):

146) 

As before the next word in P2, “where”, will be linked as a dependent of “ask”, thus disrupting the VT relation between this word and the matrix WH:

The structure is therefore ungrammatical here since there is no local source for VT to the WH element. Furthermore, the next word in P2, “they”, will have to be placed on the stack since there is nothing in P1 to license it:

The FIN element of the next word “buy” can now license “they” at the top of stack in P1, although the structure remaining in P1 will be ungrammatical for the third successive line:

Now the cell in P2 headed by FIN is ‘in sight’ of the embedded WH element of “where” in P1. In accordance with the assumptions outlined in section 5.4.5, this WH element will therefore license FIN as a complementiser in preference to the α element with which it is fused. As a result of this the parser will have to establish a new VT relation so as to allow the embedded WH to continue licensing α. Once again, though, the α element of “where” serves as an adjunct and as I have already pointed out, only a passive licensing relation will thus have to be transferred to WH. This means, in effect, that any word can be identified as a potential source of this new passive VT structure. The most local and convenient source will evidently be the FIN element itself in P2, the very element which necessitated the creation of a new VT structure in the first place:
Now that the new VT relation arising from the two elements of “where” has been resolved, it is clear that the addition of the verb “buy” in (150), with its unfulfilled licensing capacity, has also brought a local potential source for VT to the matrix WH element of “what”, which still needs to license $\alpha$:

Notice that with the processing of this tensed WH-island violation there are four ‘bad’ lines, shown in (147)-(150), as opposed to the non-finite WH-island violation above which had just two. Once again the more serious processing hiatus in the tensed example helps explain why these are consistently judged to be ‘more ungrammatical’ than corresponding non-tensed examples. Note also that taking together the discussion of the adjunct and WH islands, we find the following pattern; the violation of the non-tensed WH island in (140i) had two ungrammatical lines while the violation of the non-tensed adjunct island in (119ii) had three. Similarly the violation of the tensed WH-island violation in (140ii) had four ill-formed lines whereas the violation of the tensed adjunct island in (129ii) was associated with a five-line processing hiatus. Clearly, then, we can begin to discern here something like a hierarchy of ‘ungrammaticality’ (or, more accurately, unprocessability) associated with island constraints; I will say more about this in section 5.5.6.

5.5.4 - The relative clause island and the complex NP island

(152ii) below shows an example of a violation of the relative clause (RC) island:

It should, I think, already be intuitively obvious that relative clauses are, in effect, similar to adjuncts and that violations of the RC island will thus generally involve similar processing complexity as violations of the (tensed) adjunct island. A possible structure for (152ii) is shown below:
I assume that the processing of the first five words of (153) will proceed without any problem, yielding the grammatical structure in (154):

Here "a" has been identified as a temporary source of VT to WH. However, the next word "bar" in P2 will be linked as a dependent of "a". The resulting structure is, of course, ungrammatical, due to the absence of an alternative source for VT to WH:

Furthermore, there is nothing in P1 to license the next word, "where", which will consequently have to be placed on top of the stack:

Form-D will also be unable to link the two words which are currently 'visible', "where" at the top of the stack in P1 and "Noam" in P2. The latter will thus itself be placed on top of the stack:

The FIN element of the next word in P2, "play", can now license "Noam" at the top of the stack in P1:

---

The analysis of RCs suggested here might not be correct, however the syntactic structure of these clauses is not of immediate relevance to any processing difficulty they may cause the parser.
Note, though, that the structure at the bottom of the stack in P1 remains ungrammatical here for the fourth consecutive line. In (158) the WH element of "where", at the top of the stack in P1, is now in a position to license the cell headed by FIN in P2. In accordance with the assumptions outlined in section 5.4.5, this WH element will thus license FIN, requiring the parser to instantiate a new VT relation so as to allow it to continue licensing the α element with which it fused. Once again, the α element of "where" serves as an adjunct, and will therefore only have to be licensed passively. In this way FIN itself can be identified as a potential source of this passive VT, even though it has no unfulfilled active licensing capacity:

\[ ((([(WH _α) \ [FIN \ do]) \ you) \ know) \ a) \ bar) \ (WH _α) \ (Noam \ [FIN \ play]) \] (P1)

\[ ((([(WH _α) \ [FIN \ do]) \ you) \ know) \ a) \ bar) \ (WH _α) \ (Noam \ [FIN \ play]) \] (P2)

At this point I assume that the structure in P2 will be recognised as a relative clause and can therefore be attached as a passive dependent of the nominal "bar":

\[ ((([(WH _α) \ [FIN \ do]) \ you) \ know) \ a) \ bar) \ (WH _α) \ (Noam \ [FIN \ play]) \] (P1)

\[ ((([(WH _α) \ [FIN \ do]) \ you) \ know) \ a) \ bar) \ (WH _α) \ (Noam \ [FIN \ play]) \] (P2)

Here, then, the structure is no longer ungrammatical since "play", with an unfulfilled licensing capacity, is finally available as a local source of VT within the same cell as the matrix WH element.

\[ ((([(WH _α) \ [FIN \ do]) \ you) \ know) \ a) \ bar) \ (WH _α) \ (Noam \ [FIN \ play]) \] (P1)

\[ ((([(WH _α) \ [FIN \ do]) \ you) \ know) \ a) \ bar) \ (WH _α) \ (Noam \ [FIN \ play]) \] (P2)

Notice, though, that this (grammatical) structure is associated with a five-line processing hiatus, shown in (155)-(159), parallel to the violation of the tensed adjunct island in (129ii). This, I think, correctly reflects the relatively severe nature of RC clause violations.

There is undoubtedly more to be said about the structure of relative clauses, and the suggested analysis in (153) and (161) may not be correct. Moreover, I am unsure as to how an appropriate semantic connection could be established between a relative clause and the
nominal it 'modifies'. Nevertheless it should be clear that whatever their syntactic or semantic
structure may be, extraction from a relative clause will inevitably give rise to processing
difficulty; quite simply, being adjuncts, RCs will be passively licensed by something, and thus
extraction from them will entail a similar processing hiatus associated with violations of the
(tensed) adjunct island. More generally, the parser will only be able to administer long distance
applications of VT successfully when the dependency chain between the recipient and source
of VT contains only active licensing relations; any word interrupting this chain without an
active licensing capacity will not constitute a potential source and will thus prevent the parser
from establishing a sequence of successive VT structures, as required by the locality constraint.
This property of extraction structures can be formalised as follows:

162) *

Here B is the recipient of VT from E, allowing it to license A. However the fact that the
dependency chain between B and E contains a passive relation, between C and D, explains the
ill-formedness of this structure. This constraint alone accounts for the ill-formedness associated
with violations of the adjunct island, the co-ordinate structure island, the Relative Clause island
and, as I will show immediately below, the Complex NP island too. Of course the constraint
expressed in (162) is not part of the grammar, and is instead only a theorem of the parser's
operation in constraining VT.

Consider now the 'Complex NP' (CNP) island violation in (163i) along with its
associated structure in (163ii):

163) i. *What did you hear rumours that Sid ate _?
ii. [WH α_i] [FIN do] you hear rumours that Sid [FIN eat_i]

As I pointed out in earlier chapters, examples like this are generally felt to be only mildly
'ungrammatical', and certainly far less problematic than violations of the RC island described
above. Indeed, some speakers find certain structures analogous to (163) acceptable, and the
CNP island is even argued to be absent altogether in some languages (Horrocks 1987).

21 The constraint expressed in (162) does not, of course, account for the WH-island discussed in the
previous section.
Evidently, then, if we are to account for CNP island violations such as these in terms of the LGP's operation, the prediction should be that they should entail a fairly small processing difficulty, possibly on a par with violations of non-tensed WH islands.

However, according to the suggested representation in (163ii), CNP constructions share exactly the same structure with relative clauses. In other words, both the relative clause in (153) and the 'complement clause' in (163ii) are analysed as passive dependents of their respective nominals. The natural, though incorrect, implication of this is that violations of the complex NP island should in fact give rise to a comparable degree of 'ungrammaticality', (processing difficulty) to violations of the RC island, as illustrated by the partial processing derivation in (164):

164) i. \[ (((\text{WH}, \alpha,) \text{ [FIN do]} \text{ you}) \text{ hear}) \text{ rumours} \]

```
   P1                      P2
                               \\
```

ii. \[ ! (((\text{WH}, \alpha,) \text{ [FIN do]} \text{ you}) \text{ hear}) \text{ rumours}) \text{ that} \]

```
   P1                      P2
                               \\
```

iii. \[ ! (((\text{WH}, \alpha,) \text{ [FIN do]} \text{ you}) \text{ hear}) \text{ rumours}) \text{ that Sid} \]

```
   P1                      P2
                               \\
```

iv. \[ ! (((\text{WH}, \alpha,) \text{ [FIN do]} \text{ you}) \text{ hear}) \text{ rumours}) \text{ that Sid} \text{ [FIN eat]} \]

```
   P1                      P2
                               \\
```

v. \[ ! (((\text{WH}, \alpha,) \text{ [FIN do]} \text{ you}) \text{ hear}) \text{ rumours}) \text{ that (Sid [FIN eat])} \]

```
   P1                      P2
                               \\
```

vi. \[ ! (((\text{WH}, \alpha,) \text{ [FIN do]} \text{ you}) \text{ hear}) \text{ rumours}) \text{ (that (Sid [FIN eat]))} \]

```
   P1                      P2
                               \\
```

vii. \[ (((\text{WH}, \alpha,) \text{ [FIN do]} \text{ you}) \text{ hear}) \text{ rumours}) \text{ (that (Sid [FIN eat]))} \]

```
   P1                      P2
                               \\
```
Here then we have a similar five-line processing hiatus to that associated with the RC island violation discussed above, incorrectly implying a similar degree of ungrammaticality between the two structures.

Of course one possible way of overcoming this problem would be to reanalyse the relevant clause in (163ii) as an actively-licensed argument of "rumour" rather than a passively licensed adjunct. However, I do not believe that this is an appropriate solution for a variety of reasons. For one thing, if "rumour" were to license the embedded clause as an argument, then the implication would be that "rumour" always licensed a similar clause. This is evidently not the case however, given that the same word occurs frequently without any accompanying 'complement' clause. Moreover, if the embedded clause in (163) were a syntactic argument (active licensee) of "rumour", then extraction from this clause would be predicted to give rise to no processing difficulty at all, which is evidently not quite true; if there were a chain of active dependencies between WH and "eat" then the parser would be able to create a sequence of consecutive VT structures between the two in much the same way as it did in the grammatical examples of long-distance extraction described in section 5.4.

I believe that a more promising approach to this question might be found in the pragmatic influences that can be brought to bear in processing. Quite simply I assume that when processing (163ii) pragmatic knowledge can 'intervene' and allow the parser to attach the clause headed by "that" as a (passive) licensee of "rumour" at an earlier stage in the derivation than indicated in (164). Thus unlike the RC island violation in (153), where the entire relative clause had to be processed before it could be linked as a dependent of "bar", I assume that elements of the 'complement' clause in (163) can be linked to "rumour" before the subordinate clause itself has been fully processed. This means that any ungrammatical structure created in PI of the list can be 'rescued' much sooner by the attachment of a new local potential source of VT. I assume that what allows the parser to make this early attachment is simply the knowledge that words like "rumour" and "story" are often accompanied by a 'complement' clause. In this way although nominals like "rumour" and "story" will not actively license a clausal dependent, the presence of these clauses is still in some sense expected, if only through pragmatics rather than syntax. In this way the embedded clause in (163) could be said to occupy something of an intermediary position between an argument and an adjunct.
Consider now how (163) would be processed, taking into account this pragmatic information. Once again the processing of the first four words will be unproblematic, resulting in the same grammatical structure shown in (164i):

\[
\text{165) } \text{(([WH C T j] [FIN do]) you) hear}) \text{ rumours}
\]

The next word in P2 to be integrated into the structure in P1 is “rumours”. The Priority Principle ensures that this will be linked as a dependent of “hear”, the result being an ungrammatical structure where there is no longer a local source of VT to WH:

\[
\text{166) } \text{(((([WH g] [FIN do]) they) hear) rumours) that})
\]

Ostensibly there is nothing in (166) to suggest that a licensing relation can be established between anything in P1 and the complementiser “that” in P2. However, I assume that at this point pragmatic knowledge that the word “rumour” is often accompanied by a clause will guide the parser in allowing “that” to be attached as a passive dependent of “rumours”.

\[
\text{167) } \text{(((([WH g j [FIN do]) they) spread) rumours) t^a ;) Sid}
\]

Here, then, with the spare licensing capacity of “that”, the parser has already located a potential source of VT to the WH, rescuing the ungrammatical structure in (166). From here processing can continue normally, as illustrated by the derivation shown in (168):

\[
\text{168) i. } \text{((((([WH ttj [FIN do]) they) spread) rumours) that;) Sid [FIN eat]}
\]

\[
\text{168) ii. } \text{((((([WH ^ j [FIN do]) they) spread) rumours) that;) (Sid [FIN eat])}
\]

\[
\text{168) iii. } \text{((((([WH ctj [FIN do]) they) spread) rumours) that;) (Sid [FIN eat;])}
\]

244
The grammatical structure in (168iii) is associated with only one ill-formed line of derivation, shown in (166). In this way I believe that it is possible to reflect the fact that violations of the CNP constraint are only mildly ill-formed, especially when compared to violations of the RC island, which, as described above, involve a five-line processing hiatus.

I believe, then, that even though relative clauses and 'complement' clauses like that shown in (163) may share the same structure, these two constructions can be shown to differ in the way they are processed; it is clear that words like "rumour" and "story" are frequently accompanied by a clause, and it thus seems reasonable to suppose that this fact could be stored as part of the pragmatic knowledge relating to the behaviour of these words. This knowledge could then partially alleviate the processing difficulty that would otherwise arise from parsing a structure such as (163ii) by allowing a new local source of VT to be attached to any ungrammatical structure at a relatively early stage. Thus when "that" is linked as a dependent of "rumours" in (167), this exceptional attachment is motivated solely by pragmatic rather than syntactic properties of the noun. Generally, though, the parser would not attach a passive dependent without processing first the whole of this dependent 'phrase' (see section 5.3.3). Thus in the case of the RC island in (153), the ungrammatical structure in PL must remain as it is while the entire relative clause is processed, recognised and attached appropriately. The very nature of relative clauses means that their presence cannot be predicted from properties of the nominals that they modify. Therefore pragmatic information will not facilitate the processing of these structures.

Assuming then that the relative mildness of CNP island violations is a result of pragmatic intervention during parsing, it is interesting to note that Deane (1989) also singles out the CNP island as being particularly susceptible to the effects of such factors. He cites the following pair of examples as evidence of the pragmatic effects which may influence the acceptability of these structures:

169) i. How much money are you making the claim that the company squandered?
   ii. *How much money are you discussing the claim that the company squandered?

It is of course interesting to speculate as to the possible effects of pragmatics on other island domains, particularly the co-ordinate structure constraint (Lakoff 1986, Kuno 1987), as well as some of the other non-syntactic phenomena discussed in section 1.2.5 of the first chapter. However, the influence of pragmatics on the operation of the LGP and its consequent effects on island domains is too broad and complex an issue to be explored adequately here.
Nevertheless, I hope to have offered at least an initial insight into how pragmatic factors might be accommodated within the overall LG-based account of extraction outlined in this chapter. I certainly believe that these factors can be shown to integrate more plausibly with the parsing process than they can with a formal model of syntax, and I therefore suspect that the approach to extraction advanced here, incorporating as it does elements both from formal dependency syntax and processing, holds out the possibility of a more integrated, flexible and holistic account than would otherwise be possible within a purely syntactic framework.

5.5.5 - Predependent domains and the subject island

In section 1.3.2 of the first chapter I proposed a general and universal ban on the extraction of elements from domains which precede their head. I suggested that this generalisation allowed us to account for the absence of extraction in SOV languages like Japanese and Turkish, as well as subsuming certain cases of the subject island in English:

170)  i. Chasing sheep tires Mr. Neville.
    ii. *Sheep [chasing _ ] tires Mr. Neville.

Here the subject domain [chasing _ ] precedes its head, which will be the FIN element of the verb "tires". Consequently the extraction of the topicalised object "sheep" is ungrammatical. However, in chapter one this general ban on extraction from predependent domains lacked any principled explanation. I believe, though, that it can in fact be shown to fall out as a natural consequence of the LGP's operation. In other words, I assume that the ill-formedness of examples like (170ii) can once again be shown to arise from processing difficulty, though not necessarily the same type of processing difficulty illustrated in the other island violations examined so far.

As far as head-final languages are concerned I hope that it will already be intuitively obvious why the parser will inevitably fail to process structures involving long distance extraction; the locality constraint ensures that the parser's ability to execute a VT structure is contingent on the presence at each stage in the processing derivation of a local potential source, a head with a 'spare' licensing capacity which has so far not been linked to a dependent of its own. This means that in cases of long distance extraction a series of temporary sources

---

22Recall too the data from Modern Greek where extraction was shown to be possible both from a subject and object clause when they followed their head but impossible when these same clauses preceded their head.
must be found whose role is to ‘shuffle’ the VT relation down through the syntactic structure as it is created. However in a head final language it will evidently be impossible to come across a word with a ‘spare’ licensing capacity; by definition at the point when a word occurs it will already have to be linked to its dependents, which will have preceded it. Quite simply, then, in a SOV language nothing could ever constitute a temporary source for VT, thus explaining the absence of long distance extraction in these languages.

Returning to the example in (170ii), I believe that this can also be ruled out by reference to the LGP’s operation. More specifically, I assume that this violation of the subject island will necessarily involve illicit backtracking which causes processing difficulty and consequent ill-formedness. The first two words of this structure to be processed will be “sheep” and “chasing”, which will be placed in P1 and P2 respectively:

171)  

```
sheep [ING chase]  
P1  P2
```

Obviously the ‘intended’ reading in (170ii) involves “sheep” being linked as a dependent of “chase”. However, there is simply no way in which the parser can make this attachment grammatically here. Even though “chase” licenses a dependent, the head parameter guarantees that this must be a postdependent. Furthermore, VT is inapplicable here since the only possible recipient of “chase”’s valency, the ING element, is also unable to license a predependent. Thus the operation of Form-D will fail here and the processor’s only option is to place “chasing” above “sheep” in the stack:

172)  

```
sheep [ING chase]  [FIN tire]  
P1  P2
```

The FIN element of the next word, “tires” is of course capable of licensing a predependent. However, the only word in P1 which is visible to the processor and thus capable of being licensed by FIN is “chasing”, at the top of the stack:

173)  

```
sheep ([ING chase] [FIN tire])  
P1  P2
```

Now the parser can allow FIN to license the headless nominal “sheep” in P1 through VT.
There are, of course, already two potential local sources of VT in P2, the verbs “chase” and “tire”. Following the discussion in section 5.4.4, however, it should be clear that “tire”, a direct dependent of the VT recipient, FIN, will be a more local source of VT to FIN than “chase” is. For this reason “tire” rather than “chase” will automatically be identified as the source of VT:

![Diagram](image)

Here, then, the licensing capacity of the verb “chase” remains unfulfilled. Furthermore, there is no longer any way in which this verb could grammatically license a dependent of its own; given that “chase” occurs in an embedded cell, marked A in (174), any attempt to link this verb as a head of a dependent in P2 would, in effect, constitute a backtracking operation bypassing the outermost cell of the structure, marked B. As I pointed out in section 5.3.5, this will give rise to processing difficulty since backtracking is predicted to be restricted to the outermost cell only. Thus in (174) the only way in which “Neville” in P2 can be linked to the structure in P1 is as a dependent of “tire”, through the Priority Principle:

![Diagram](image)

The structure here is ungrammatical since there is no local source of VT to FIN, meaning that this element will only be able to license “sheep” passively, like an adjunct.

Unlike all the other island violations discussed above, in (175) the parser is confronted with an ungrammatical structure at the last point in the processing derivation, with nothing left in P2 to work on. In the other island violations seen so far the final structure was always grammatical, even though it had been ungrammatical at certain stages in its processing history. (This of course accounted for the ill-formedness of these examples). In (175), however, the only option left to the parser in seeking to create a grammatical structure is to employ backtracking and reanalysis. Unlike a simple ungrammatical structure, such as (81)/(83) above, backtracking is a possible option here since the dependent “sheep”, which is not properly licensed, is matched with a head, “chase” with an unfulfilled licensing capacity. However the reanalysis required to link these two elements through VT involves two elements, the verb “chase” and FIN, which occur in the furthest embedded cell, A. Thus even if it were possible to arrive at the grammatical structure in (176) from (175), the parser would only be able to do...
so at an unacceptable processing cost. In this way the subject island violation in (170ii) will either be ungrammatical, as in (175), or virtually unprocessable:

176)

The above reasoning applies equally to other cases of the subject island, such as (177ii) below where the bracketed subject domain follows its head:

177)  
  i. [Eating beetles] causes heartburn.
  ii. *What does [eating _]cause heartburn?

A processing derivation for (177ii) is given in (178):

178)  
  i. 

  ii. 

  iii. 

  iv. 

  v. 

In (178v) the parser is once again left with an ungrammatical structure, there being no suitable local source of VT to the WH element allowing it to license \( \alpha \) actively. Of course one option for the parser here would be to backtrack and reanalyse “eat” as a source of VT to WH, assuming that the parser has not already discounted a transitive reading of this verb. However, “eat” occurs in the embedded cell marked B, whereas WH, the supposed recipient of VT
occurs in the furthest embedded cell of all, A. For this reason any attempt to establish "eat" as a new source of VT to WH in (178v) would be tantamount to illicit backtracking and reanalysis across two external cells, D and C. This of course is predicted to give rise to unacceptable processing difficulty.

In this way just as with the island violation in (170ii)/(175) the parser is faced with an intractable problem in (178v); it can either leave the ungrammatical structure as it is, or it can seek to establish a grammatical alternative, resulting in processing effort and consequent ill-formedness. Notice, though, that the processing difficulty associated with these subject island violations, arising from illicit backtracking, is different from the processing difficulty that was shown to characterise the other island violations discussed in previous sections. This, I believe, may account for the fact that in many ways the subject island behaves differently from other islands in English (see section 1.3.2 of the first chapter). For example, the non-tensed examples of the subject island in (170ii) and (177ii) are notably worse than non-tensed examples of other islands. Compare, for example, (170ii) and (177ii) with the violation of the non-tensed adjunct island in (119ii) and the WH island in (140i) above.

5.5.6 - Summary of the island data

In the previous sections I suggested that island constraints can be shown to arise from limitations on the parser's ability to process certain types of structure. In this way violations of these islands can be described in terms of processing difficulty rather than ungrammaticality. I have suggested that there are three basic advantages to this view of extraction constraints. Firstly, 'processing difficulty', in being a relative term, implies a more graded approach which better suits the scalar nature of island violations, particularly the variation in acceptability between tensed and non-tensed examples. Secondly, as I noted in section 5.5.4, there is good reason to think that processing is more amenable than syntax to the influence of non-syntactic factors such as semantic and pragmatic criteria. In this way an account of extraction based partly on processing has, in my opinion, a more realistic chance of offering a holistic and integrated account of a wider range of phenomena than a purely syntactic account. Finally, if extraction constraints can be shown to arise from incidental properties of the parser's operation, as I have suggested, then this allows the syntactic underpinnings of the account to remain straightforward and elegant. In this way the simple VT rule stated in (68) can be left as it is, with the intricacies of extraction constraints falling out independently from the way in which VT is executed 'on line' by the LGP.

250
As I have just shown, one way in which the parser can experience difficulty in processing an extraction structure is in being forced to carry out illicit backtracking and reanalysis of structure contained within an embedded cell. In section 5.3.5 I suggested that backtracking should be restricted to the external cell of a structure at any given time in its processing derivation, and thus reanalysis of elements embedded inside the external cell is predicted to be problematic for the LGP. This type of processing difficulty defines one type of island domain, exemplified in English by the subject island. The processing difficulty associated with violations of the other island domains (adjunct, co-ordinate, WH, relative and ‘complex NP’ clauses) is different, and can be shown to arise from the parser’s inability to avoid creating a temporarily ungrammatical structure during processing. Thus while the final output of the parsing process in these examples may be grammatical, in the sense that it is generated by the lexicon and head parameter, it will be ill-formed on account of there having been one or more stages of processing hiatus in its derivation at which the structure was ungrammatical.

The important point about this latter type of island constraint is that it is possible to compute varying degrees of processing difficulty, and consequent ill-formedness, on the basis of the number of stages (or lines) in the derivation at which the structure was ungrammatical. A structure which is ungrammatical at four stages in its processing derivation will be more difficult to process, and thus more ill-formed, than a structure with just two such problematic lines. In this way it is possible to establish a hierarchy of ill-formedness according to the size of the processing hiatus associated with each island. This hierarchy is illustrated in (179) below where the numbers in parenthesis refer to the processing derivations shown in sections 5.5.2, 5.5.3 and 5.5.4.

179) i. One ungrammatical stage: Complex NP island (165) - (168iii)
   ii. Two ungrammatical stages: WH island (non-tensed) (141) - (145)
   iii. Three ungrammatical stages: Adjunct island (non-tensed) (120) - (128)
       Co-ordinate structure island (non-tensed) (139i-v)
   iv. Four ungrammatical stages: WH island (tensed) (146) - (151)
   v. Five ungrammatical stages: Adjunct island (tensed) (130) - (137)
       Relative clause island (154) - (161)

In this way it is possible to envisage a continuum of well-formedness ranging from fully ‘grammatical’ examples of long distance extraction such as those discussed in sections 5.4.4
and 5.4.6, which have no problematic stages in their processing derivation, through to totally ill-formed examples which involve a four or five line processing hiatus. Less severe examples of island violations will occupy intermediate positions at different points of this spectrum. In this way I believe that it is possible to offer a more flexible and accurate account of extraction constraints than would otherwise be possible within a purely syntactic account which implied an absolute distinction between grammaticality and ungrammaticality.

5.6 - Extraction Asymmetries

5.6.1 - Objects v adjuncts

So far all examples of island violations have involved the illicit extraction of an object. This has been useful since it allows us a fair comparison of the relevant island domains without having to take into account the various interfering effects arising from asymmetries in the extractibility of different elements. Nevertheless, as I have already pointed out in previous chapters, adjuncts do consistently give rise to a more severe degree of ill-formedness when extracted from an island than objects extracted from the same domain, a contrast illustrated in (180) below:

180) i. The neighbours complain [if Noam plays the banjo on Tuesdays].
    ii. *What do they complain [if Noam plays _ on Tuesdays]?
    iii. **When do they complain [if Noam plays the banjo _ ]?

Whereas (180ii) is ill-formed, (180iii) would appear to be totally uninterpretable under the intended reading. As I noted in section 1.2.6, this asymmetry between objects and adjuncts applies equally to all islands, and in the second chapter I described how in order to account for this contrast PP Theory invokes some of its most complex and intricate syntactic machinery in the form of the ECP and its related notion of proper government. Recall too from the third chapter that WG was unable to offer any account for this important asymmetry between arguments and adjuncts.

However, I believe that it is possible to explain the contrast illustrated in (180) simply and at no extra cost by careful consideration of some of the aspects of LG and its related parsing system already discussed. Consider first of all the two grammatical sentences and their syntactic/semantic representations in (181) and (182) below:
181) i. What did Wally wear?
   
   ![Diagram](image1)

   ii. Why did he wear it?

   ![Diagram](image2)

As I noted in section 5.4.5, the two $\alpha$-elements in (181ii) and (182ii) behave very differently both in terms of syntax and semantics, as illustrated by their respective representations; the $\alpha$-element in (181ii), in common with other arguments, is actively licensed (through VT) and assigned a semantic role by another word in the sentence, "wear". The $\alpha$-element of "why" in (182ii), though, like other adjuncts, is licensed passively, again through VT, and itself assigns a semantic (Subject) role to "wear".

I believe that this difference in the behaviour of the two $\alpha$ elements in (181) and (182) is reflected in how the two structures are processed; I have already described in some detail how when processing a sentence such as (181) the parser will have to locate a word with an unfulfilled licensing capacity to act as a source of VT to WH, allowing this head to continue licensing $\alpha$ actively. What the parser must therefore identify in cases of argument extraction such as (181) is a VT-source like "wear" with a specific syntactic property, namely a 'spare' valency. Consider now, though, how (182) would be processed; the first two words to be recognised are "why" and "did", placed in P1 and P2 of the list respectively:

183) ![Diagram](image3)

Assuming that the presence of FIN in P2 will activate the complementiser function of WH in P1, the latter will license FIN in preference to the $\alpha$ element with which it is fused:

184) ![Diagram](image4)
In order to maintain the internal structural integrity of the word "why", the WH complementiser element will have to continue licensing $\alpha$. Since the $\alpha$ element of "why" will function as an adjunct, though, I assume that it will only have to be licensed passively like other adjuncts\(^{23}\). Initially, then, there would appear to be no reason why the WH element of the same word could not continue to license $\alpha$ passively while also licensing the FIN element of "did" actively:

\[ (\text{WH} \quad \alpha) \quad \text{[FIN do]} \quad \text{he} \]

\[ \text{P1} \quad \text{P2} \]

However, bearing in mind the discussion of adjuncts in section 4.4.6 of the previous chapter, it should already be clear why the structure in (185) is incompatible with the sentence in (182i); being an adjunct the $\alpha$ element of "why" will assign a Subject role to whatever it serves to 'modify'. According to the principle of Subject identification stated in (65) of section 4.4.3, a predicate will assign its Subject role to its closest d-commander. This means that in (185) $\alpha$ could only assign its Subject role to, and thus 'modify', FIN, an unlikely situation given that adjuncts invariably modify verbs or nouns. Moreover, by allowing WH to license $\alpha$ passively in (185) the parser is, in effect, committing $\alpha$ to modify FIN far too early, before any other structure has been processed. As I noted in section 5.3.3, the syntactic head of a passive licensee (adjunct) will be determined by semantic properties of the licensee itself. There is therefore no point in linking an adjunct to a head until such time as this attachment is appropriately motivated by the semantic requirements of the adjunct.

The parser's only way of avoiding this problem posed by (185) is to initiate a VT structure at (184). This will allow WH to continue licensing $\alpha$, with a passive licensing capacity acquired through VT, while also delaying any commitment as to the semantic argument structure of the adjunct element $\alpha$; the ultimate source of any VT to WH in (184), rather than WH itself, will determine the location of $\alpha$'s closest d-commander, and since this VT-source need not be found until later on, the parser will thus be able to hedge its bets temporarily as to what "why" will modify. Let's assume then that at (184) the parser will proceed with a VT structure, suitably motivated by the need to satisfy the semantic requirements of the WH.

---

\(^{23}\)It is true that in (183), prior to being disconnected from its WH head, the $\alpha$ element was licensed actively like an argument. As I noted in section 5.4.5 though, at this point the element in question does not actually function as an adjunct, and is merely part of a self contained WH-word, "how".
requirements of $\alpha$. Since $\alpha$ is an adjunct, though, all that will have to be transferred to WH to allow it to continue licensing $\alpha$ is a passive, and thus nebulous, licensing capacity. Any word is capable of licensing a passive dependent, and therefore any subsequent word could trivially serve as a source of this passive VT to WH. However, the most local candidate in (184) is FIN, WH's direct dependent, and consequently this will be identified as the source of VT:

186) \[\begin{array}{c}
\text{([WH $\alpha$,] [FIN do,])} \\
P1 \quad P2 \\
\text{he}
\end{array}\]

Here then the WH element of "why" continues to license $\alpha$ passively, with an 'invisible' valency acquired from FIN through VT. In this way, as discussed in section 5.4.1, FIN could, in a sense, be described as $\alpha$'s true head. FIN's other dependent, "do" will therefore be $\alpha$'s nearest d-commander in (186). Once again, though, the auxiliary verb "do", like FIN, is an unlikely recipient of $\alpha$'s Subject role and thus I assume that a new VT structure will have to be created so as to allow $\alpha$ to be d-commanded by a more suitable recipient of its Subject role. Of course this new VT relation will have to obey the locality constraint, and the only other word in the cell in (186) which could serve as a new source of VT to WH is "do":

187) \[\begin{array}{c}
\text{([WH $\alpha$,] [FIN do,])} \\
P1 \quad P2 \\
\text{he}
\end{array}\]

However, "do", having no dependent of its own yet, does not provide a new d-commander for $\alpha$, and its Subject role will thus remain unassigned. However, "he" in P2 can be linked as a dependent of FIN, the only word in P1 capable of licensing a nominal:

188) \[\begin{array}{c}
\text{([WH $\alpha$,] [FIN do,])} \\
P1 \quad P2 \\
\text{he} \quad \text{wear}
\end{array}\]

Similarly "wear" in P2 can now be attached as a dependent of the auxiliary verb "do":

189) \[\begin{array}{c}
\text{([WH $\alpha$,] [FIN do,])} \\
P1 \quad P2 \\
\text{he} \quad \text{wear} \quad \text{it}
\end{array}\]
Even though "do" has been linked to a dependent of its own here, it can still function as a source of VT since all that is transferred to WH is a passive licensing capacity which has no bearing on its ability to license its own dependent actively. However, with the attachment of the verb "wear" the α element of "why" has found a plausible recipient of its Subject role. Of course at this point the incremental parser will not necessarily know whether α will actually modify "wear"; the structure in PI of (189) is not complete since "wear" itself still has an unfulfilled licensing capacity, implying that there will be more structure to be processed. Nevertheless, there is no reason why in (189) α should not assign its Subject role to "wear" provisionally, in the absence of any more plausible 'modifyee'. Moreover, the verb "wear" is already the closest d-commander to α, since "do", its head, is the source of the licensing relation linking the WH and α elements of "why":

Here then "wear", being the closest d-commander to α, will receive its Subject role, though only provisionally, hence the dotted arrow marked S in (190). Processing can now continue, with "it" in P2 being linked as a dependent of "wear" in P1:

At this point the structure in P1 could be described as being syntactically complete since the licensing properties of all the words contained within it are satisfied, and there is nothing left in P2 to be processed. For this reason the parser can safely assume that the provisional Subject role assigned in (190), along with its resulting syntactic VT structure, was correct. Thus the α element of "why" will 'modify' the verb "wear", as already illustrated in (182ii) above.

The above example serves to illustrate how the processing of a VT structure involving adjunct extraction differs radically from the processing of a VT structure involving the extraction of an argument. Whereas in the latter case a source of VT is sought on the basis of its syntactic capacity to provide a spare licensing valency, in cases of adjunct extraction the source of VT will instead be identified according to the semantic requirements of the extracted
adjunct itself, the VT-dependent α in (191). In other words, what determines that “do” in (190/191) is the source of VT to WH is not any specific syntactic property of “do” itself, but rather the fact that it allows a d-command relation to be maintained between α and the verb “wear”, which α modifies. Here, then, just as with the attachment of the adjunct in 5.3.3, we can see again how in creating syntactic structure the LGP will be susceptible to the influence of semantics as well as pragmatics (Rayner et al. 1983, Mitchell and Holmes 1985).

Exactly the same will also be true in cases of long distance WH-adjunct extraction such as (192) below:

192) Why did they say [he wore it _ ]?

After processing the first four words, the parser will recognise “say” as a possible semantic argument of the adjunct α element of “why”. In the absence of a more likely candidate the parser will allow α to assign its Subject role provisionally to “say”. Once again this means that “say” must be allowed to d-command α so as to receive its Subject role. In this way the immediate head of “say”, the auxiliary verb “do”, will have to be identified as the source of a passive VT to the WH element of “why”:

193)

Once again, though, the structure in P1 is incomplete here since the unfulfilled licensing capacity of the verb “say” suggests that there is more input to be processed. For this reason the Subject role assigned to this verb will have to remain provisional in case another recipient of this role is found later. Further processing will in fact reveal that “say” licenses an embedded clause containing the verb “wear”, another possible Subject for α. I assume, then, that the parser may allow another provisional assignment of this role to “wear”:

194)

Of course this entails that, temporarily at least, the passive VT to WH will have two provisional sources, “do” and the embedded FIN; these allow α to be d-commanded by, and
thus assign a Subject role to, "say" and "wear" respectively. Nevertheless, the α element of "why" will only assign one Subject role, and thus one of the provisional roles in (194), along with its associated VT structure, will have to be eliminated. Again it will be up to the semantics to decide which of the verbs will be modified by "why", and this will in turn determine the true syntactic source of the passive VT relation to WH. Assuming a 'long distance' interpretation in (192) whereby the question being asked concerns the reason for wearing something, α's Subject role will in fact be assigned to "wear", and thus the provisional role assigned to "say" in (193)/(194), and the consequent VT structure between "do" and WH, will be cancelled, leaving just one semantic role and syntactic VT structure:

It may initially appear that the elimination of the provisional Subject role assigned to "say" in (193/194) and its resulting VT structure may equate to illicit backtracking into an embedded cell. In fact this is not really this case, though, since no real structural commitment was made in (193); given that in purely syntactic terms any word could qualify as the source of a passive VT, by establishing "do" as a source of VT the parser was really only biding its time, allowing α to assign its Subject role provisionally, a decision which, by its very nature, could be undone at a later time.

Once again, then, in (192)-(195) we see how in cases of adjunct extraction the operation of VT will be governed almost entirely by semantic considerations. Unlike cases of WH-argument extraction, such as those discussed in section 5.4, where the parser must locate a word with a spare licensing capacity to satisfy the syntactic requirements of the extracted argument, in cases of adjunct extraction the parser will instead seek a VT-source which allows the semantic role of the adjunct to be assigned appropriately. Thus, although in the above processing derivation the source of VT is 'passed down' from "do" to the embedded FIN, rather like cases of argument extraction, the reasons behind this are entirely different, and concern the semantic properties of α alone.

^I will say more about structures containing two or more VT sources in section 5.6.3.
I believe that this distinction between the processing of VT structures involving arguments and adjuncts can be shown to account for the extraction asymmetries illustrated in (180); consider once again, for example, the sentence in (180iii), repeated below as (196):

196) **When do they complain [if Noam plays the banjo _ ]?**

As I noted above this example, under the intended reading, is uninterpretable, as opposed to (180ii) which is interpretable but ill-formed. After processing the first four words of (196) the parser will yield the grammatical structure in (197):

197) 

Here the verb "complain" will be recognised as a likely candidate to be modified by the α element of "when". Furthermore, the VT structure between "do" and WH ensures that the requisite d-command relation already exists between α and "complain" allowing the former's Subject role to be assigned to the latter. However, unlike (193) above, here the structure in P1 is syntactically complete, in the sense that the licensing properties of all the words contained within it are satisfied. As far as the structure in P1 is concerned, then, there is no indication that any alternative recipient of α's Subject role will be found, and consequently this role will be assigned to "complain" definitively rather than provisionally:

198) 

It is true that in (197)(198) P2 contains the word "if", suggesting that there will be more structure to be added to P1, possibly containing another potential recipient of α’s Subject role. However, bearing in mind the discussion of adjunct attachment in section 5.3.3, it is by no means clear at this stage when, or even if, the parser will be able to attach "if" grammatically to anything in P1; certainly no such attachment will take place before "if" has been linked to its own dependent. In (198) α’s Subject role is thus assigned definitively on the basis of the complete and well-formed structure in P1. The rest of the sentence will be processed as follows:
Notice, then, that none of the stages in the derivation in (199) are ungrammatical. Neither, however, are any of these stages associated with the interpretation intended in (196).

What we are left with in (199vii) is a grammatical structure, associated with no processing difficulty, in which the $\alpha$ element of "when" modifies the verb "complain" rather than "play".

It is true that when "play" is linked to the structure in P1 in (199v) it could, theoretically, be a possible recipient of $\alpha$'s Subject role. However, the definitive assignment of $\alpha$'s Subject role to "complain" in (198), along with the syntactic VT structure which made this possible, means that no other assignment of this role is possible in (199). Any attempt to go back and cancel the original Subject role and VT relation would constitute illicit reanalysis of syntactic
structure contained in an embedded cell. Furthermore, since the structure in (199vii) is perfectly grammatical, there is nothing to motivate any such backtracking in the first place.

This, I think, explains the 'uninterpretability' of the example in (196) with its intended reading, there is simply no point trying to process into the embedded island clause seeking a new source of passive VT to WH, since a source is already available in the matrix clause associated with a perfectly acceptable interpretation. In this way a reading whereby "when" modifies "play" is unobtainable, hence the uninterpretability of (196). Of course in cases of ill-formed argument extraction, such as those discussed in section 5.5, the parser is forced to continue seeking a VT-source within the island clause since what is required is a word with an identifiable syntactic property, a spare licensing capacity. Thus the VT-source in examples of argument extraction could be said to have a recognisable syntactic 'tag' which ensures its recognition, even within an island domain. So, for example, in (180ii) above the only possible source of the active VT to the WH element of "what" is "play", since this is the only word in the structure with a spare licensing capacity. The parser will therefore have to derive the intended interpretation in (180ii), even though it results in processing difficulty and consequent ill-formedness. In this way I believe that it is possible to account for the 'ungrammaticality' of (180ii) as opposed to the uninterpretability of (180iii).

The same extraction asymmetry will apply equally to all the island domains discussed in section 5.5, in each case the parser will assign the Subject role of the 'extracted' α adjunct definitively, rather than provisionally, to the first potential recipient encountered in the matrix clause, thus precluding the possibility of this role's assignment to anything in the embedded island clause. This is because in each of the island constructions, at the point where α's Subject role is assigned, the matrix syntactic structure will be complete, in that it contains no word with an unfulfilled licensing capacity, the implication of this being that no other recipient of the Subject role will be found. Indeed it is precisely this 'syntactic completeness' of the matrix clause which defines adjunct, WH, relative and complex NP subordinate clauses as islands for argument extraction too; for example in the (non-tensed) WH-island violation in (140i) of section 5.5.3 above, the fact that the structure in P1 has no spare licensing capacity in (142) means that there is no source available there for an active VT relation with the WH element of "what". This causes a processing hiatus which is only resolved in (144) where a new

25 Recall from section 5.5.5 that, unlike here, in cases of argument extraction from a subject island illicit backtracking was only way of 'rescuing' an ungrammatical structure.
potential source, in the form of "to", is attached to the structure. If, therefore, in (142) the \( \alpha \) element of the WH-word had instead been an adjunct which assigned a semantic role, rather than an argument which receives one, then this role would have to be assigned to "ask" definitively, under the (mistaken) assumption that no other recipient would be found later:

\[
(((\text{WH } \alpha)) \rightarrow \text{ask}) \rightarrow \text{to} \]

In this way an interpretation whereby the \( \alpha \) element of an extracted WH-adjunct 'modified' a verb such as "buy" in the embedded WH-clause would be unobtainable since it would involve backtracking and eliminating the original (definitive) semantic role and syntactic VT structure illustrated in (200). Once again, then, I believe that this accounts for the argument/adjunct asymmetry below where (201i) is (mildly) ill-formed whereas (201ii), displaying long distance extraction of the WH-adjunct from the embedded clause, is uninterpretable:

201) i. (?)What did Wally ask [where they bought _ ]?
ii. **Where did Wally ask [what they bought _ ]?

What I hope to have shown here, then, is how islands for argument extraction will, by definition, also be islands for adjunct extraction. The increased 'ungrammaticality' associated with the latter case can be shown to arise from the more flexible, semantics-oriented nature of the 'passive' VT involved in cases of adjunct extraction; very simply, my suggestion is that since a well-formed VT structure (and associated semantic interpretation) can inevitably be established between a displaced adjunct and the matrix clause, there will be no motivation for the parser to expend the processing effort needed to locate another VT-source within an embedded island clause. This of course contrasts with the situation with extracted arguments where the parser will specifically require a word with a spare active licensing capacity as a VT-source, even if it is located in an island domain. In this way the apparently complex extraction asymmetries between arguments and adjuncts can in fact be attributed to the relatively simple

\[^{26}\text{Notice that there is no real distinction between the ill-formedness of the example in (201ii) here and the (tensed) adjunct island violation in (196). In fact extraction of an adjunct seems to be equally bad from all island domains, suggesting that the argument/adjunct asymmetry overrides the scalar effects of island domains, illustrated in (179) above. Thus even adjunct extraction from a relatively 'mild' island, such as the non-tensed WH-island in (201ii), is significantly worse than extraction of an object from a much stronger island, such as the tensed adjunct island in (129) of section 5.5.2.}\]

262
The distinction between active and passive licensing, discussed in section 4.3.6 of the fourth chapter.

5.6.2 - Subjects v objects

The pervasive extraction asymmetry between adjuncts and objects discussed in the previous section is not the only factor which interacts with the effects of the island constraints; in the first three chapters I noted a similar contrast concerning the 'extractability' of subjects and objects, illustrated in (202) and (203) below:

202) i. Willy would laugh [if Wally wore a fez].
   ii. *What would Willy laugh [if Wally wore _ ]?
   iii. **Who would Willy laugh [if _ wore a fez]?

203) i. Willy knew [why Wally likes hats].
   ii. *What does Willy know [why Wally likes _ ]?
   iii. **Who does Willy know [why _ likes hats]?

Here, then, extraction of a subject from the bracketed island clauses is consistently worse than extraction of an object.

In section 1.3.3 of the first chapter, I suggested that this asymmetry between object and subject extraction could be accounted for simply by extending the 'that-trace filter', the effects of which are shown in (204):

204) i. Sid said [(that) his granny had seen 'Eraserhead'].
   ii. Who did Sid say [ _ had seen 'Eraserhead']?
   iii. *Who did Sid say [that _ had seen 'Eraserhead']?

In older versions of PP Theory the example in (204iii) would be ruled out because the subject 'gap' occurs next to the complementiser "that", in violation of the 'that-trace' filter whose specific function is to ban this configuration (Chomsky and Lasnik 1977)\(^\text{27}\). However, in section 1.3.3 of the first chapter I argued that it might be possible to extend the 'that-trace' filter to a more general 'Complementiser Gap Filter' (CGF) which prevents extraction of any element which occurs adjacent to a complementiser. If, then, we take words like "if" in (202) and "why" in (203) to be a form of complemetiser, as well as "that" in (204), then the

\(^{27}\text{My reference to 'gap' here is purely for the sake of convenience, and refers only to the position in which a word would occur had it not been extracted.}\)
additional degree of ungrammaticality of the examples in (202iii) and (203iii) over those in (202ii) and (203ii) could thus be explained because these sentences violate the CGF as well as the relevant island constraint. Of course (204iii) only violates the CGF and no island constraint.

Needless to say, the CGF does not affect subjects extracted from clauses which are not introduced by a complementiser, as illustrated by the example in (204ii) above. Similarly, as I noted in the first chapter, the CGF, like the ‘that-trace’ filter, only applies in cases where a subject and complementiser are immediately adjacent. Extraction is thus unproblematic where the complementiser and subject ‘gap’ are separated by other material:

205) Who do you think [that Sid said [ _ had seen ‘Eraserhead’]]?

So too the CGF correctly predicts that extraction of a subject from an island clause, where the relevant ‘gap’ is not adjacent to a complementiser, should be no worse than an object:

206) i. They were amazed [when Sid said his granny had seen ‘Eraserhead’]

   ii. *What were they amazed [when Sid said his granny had seen _ ]?

   iii. *Who were we amazed [when Sid said _ had seen ‘Eraserhead’]?

According to my intuitions, the examples in (206ii) and (206iii) are fairly consistent with respect to their ill-formedness, and do not display an analogous asymmetry to the sentences in (202) and (203) above. Moreover, recall from section 2.5 of the second chapter that PP Theory approach to this question, based on inherent structural asymmetries between subjects and objects, was unable to account for this type of example.

I believe that the CGF is a valuable part of LG, and allows a very much simpler account of the otherwise troublesome asymmetry in (202) and (203), as well as subsuming the well-known effects of the ‘that-trace’ filter illustrated in (204). However what we still lack is any principled explanation for the CGF; why should it be that extraction of a subject which is adjacent to a complementiser gives rise to ungrammaticality? I do not have any fully worked out or definitive answer to this question, and what follows is only a tentative, somewhat speculative attempt to reduce the CGF to more primitive properties of the syntax.

In section 4.3.1 of the previous chapter I suggested that complementisers could be defined as nominals which license FIN, see (207i) below. In this way they can be distinguished from determiners, nominals which license another nominal, (207ii), and substantives, which license nothing (207iii):
However, one of the reasons why I first suggested that conjunctions like "if", "before" and "and", as well as WH-words, might be classed as complementisers is that they all serve to introduce a clause:

208) i. He said [that she saw it]
   ii. We know [what Santa stole]28
   iii. He threw up [after he ate it]
   iv. Balanescu played 'Swan Rot' [and his neighbours rang the police]  

In my opinion, a complementiser’s specific function of introducing a clause is not adequately reflected in its lexical entry, indicated in (207i). It is true that clauses are headed by FIN, the licensee in (207i)29. However, a clause is actually a constituent comprising the direct and indirect dependents of FIN as well as FIN itself, and it might thus be difficult to allow a complementiser to license a clause as a syntactic entity without also countenancing some form of primary constituency. However, I believe that there is another possible solution to this problem; given that a clause will correspond roughly to the dependency structure of FIN, one way in which a complementiser may come close to selecting a clause as a dependent would be to license a saturated FIN (FIN\textsubscript{Sat}), a FIN element which had already been linked to all its dependents. This requirement can be stated simply as a lexical property of complementisers by modifying the entry in (207i) as follows:

209) \[ \text{N} \quad \text{FIN}_{\text{Sat}} \]

Bearing in mind, then, that a FIN element will license a verb (with which it is fused) and a nominal, the subject of the sentence, the implication of (209) is that these dependents must be linked to FIN, thus ‘saturating’ it, before FIN itself can be linked as a dependent of a complementiser. In this way, unlike other words, a complementiser will be able to guarantee that instead of licensing a single element such as FIN, its dependent will actually be overtly ‘clausal’ in nature, consisting minimally of FIN and its attached verb and subject dependents.

---

28WH-words are exceptional in being able to introduce both a matrix and embedded clause.

29Recall that in section 5.5.3 I suggested that the non-finite marker "to" might be represented as FIN\textsuperscript{o}, implying that even non-finite clauses will be headed by a FIN element of sorts.
I have already referred to the ‘saturation’ of a head in the locality principle, stated in (71) of section 5.4.2, which constrains the application of VT. Just as with that case, the saturation of the FIN element in (209) fixes it in a specific point in time, after it has been linked with its dependents. This again naturally points to the operation of the parser, which is of course responsible for the incremental fulfilment of words’ licensing properties. I assume then that the LGP will play a crucial role in ensuring that the somewhat unusual lexical requirement in (209), with its diachronic connotations, is satisfied. Consider first of all how the grammatical sentence in (208i) above would be processed. Parsing the first three words will yield the structure shown in P1 of (210) below:

\[
\begin{array}{c}
\text{((he}} \quad \text{[FIN say]}) \quad \text{(that)} \quad \text{she} \\
P1 \\
\end{array}
\]

The next word in P2, “she” cannot be linked to anything in P1 and will thus be placed on top of the stack:

\[
\begin{array}{c}
\text{((he}} \quad \text{[FIN say]}) \quad \text{(that)} \quad \text{she} \quad \text{[FIN see]} \\
P1 \\
\end{array}
\]

Now “she” at the top of the stack in P1 can be attached as a dependent of FIN in P2, creating a new cell there:

\[
\begin{array}{c}
\text{((he}} \quad \text{[FIN say]}) \quad \text{(that)} \quad \text{(she}} \quad \text{[FIN see]}) \\
P1 \\
\end{array}
\]

At this point FIN, the head of the cell in P2, can be licensed by the complementiser “that” in P1. Notice that in accordance with the lexical entry in (209) FIN is saturated, and has already been linked to its two dependents “she” and “see”. Indeed, the licensing relation between FIN and its associated verb will inevitably ‘predate’ any other syntactic attachment since the two are amalgamated in a single word (see section 5.3.5). For this reason the ‘clausal’ cell in P2, headed by FIN, can be attached as a dependent of “that”:

\[
\begin{array}{c}
\text{((he}} \quad \text{[FIN say]}) \quad \text{(that)} \quad \text{(she}} \quad \text{[FIN see]}) \quad \text{it} \\
P1 \\
\end{array}
\]
Finally, Form-D will link “it” in P2 as a dependent of “see”:

214)

Here we can see how in a well-formed structure a complementiser such as “that” will license a saturated FIN, in accordance with its lexical requirements. The same will also be true of other complementisers such as conjunctions and WH-words (see, for example, the discussion of the sentence in (140ii) in section 5.5.3). However, (215) below shows an example of a simple CGF-violation:

215) *Who did he say that _ saw it?

The first five words of this sentence will be processed without any difficulty:

216)

At this point “that” in P1 should license the FIN element in P2. Here, though, the lexical property of the complementiser, stated in (209), is not satisfied since FIN still lacks one dependent and is thus not saturated. Nevertheless, I assume that since FIN is at least the right category to be licensed by “that”, the parser will have to make the necessary attachment, even though it results in ungrammaticality:

217)

Notice that here the embedded FIN has replaced “that” as the source of VT to WH, in accordance with the principles of VT discussed in section 5.4. One result of this is of course that in (217) this FIN element has found a second dependent, α, and is thus saturated. However, this saturation only occurs after (and as a result of) FIN’s attachment to “that”, and it thus comes too late to prevent the (partial) violation of the complementiser’s lexical requirement that it be linked to a FIN element that is already saturated.$$^{30}$$

30Of course had the complementiser been absent in (215) then in (216) the unsaturated FIN element in P2 would have been linked as a direct dependent of the matrix verb “say”. This would not cause
Finally, "it" in P2 of (217) will be linked as a dependent of "see":

\[
(((\text{[WH } \alpha, ] \text{ [FIN do] he say) that}) \text{ [FIN, see]) it})
\]

Although the final structure here is grammatical, in that the licensing properties of all its constituent words are satisfied, the processing derivation leading to this structure was associated with the problematic attachment of the unsaturated FIN element with "that", in violation of the latter's licensing requirements. Here, then, once again we can see how the global operation of the parser in processing a structure will play a crucial role in determining its well-formedness or ill-formedness.

What I have tried to show here, though, is how the CGF constraint can follow from a specific lexical property of complementisers, stated in (209). In this way the complex subject/object asymmetries illustrated in (202), (203) and (204) above can be shown to arise from the fact that complementisers only license a saturated FIN. Thus, for example, the illicit extraction of the subject in (203iii) not only violates the WH-island, but also means that the WH-complementiser "why" can only license FIN before it has been saturated. This accounts for the more severe degree of ill-formedness associated with (203iii) over (203ii). It is true that as it stands the generalised lexical entry for complementisers in (209) is itself a stipulation. However, all lexical entries are to some extent stipulative, and the lexicon is therefore a more plausible repository of specific information than \textit{sui generis} principles of the grammar such as the CGF. Moreover, I believe that this approach, in reducing extraction asymmetries to lexical properties of complementisers rather than subjects or objects themselves, is better suited to a dependency formalism such as LG which, unlike PP Theory, does not recognise inherent configurational distinctions between subjects and objects.

5.6.3 - A final note on parasitic gaps

One issue which has so far not been addressed in this chapter is the question of 'parasitic' or multiple gaps (Chomsky 1982, Engdahl 1983, Hudson 1984-b). (219ii) offers an example of a relevant construction:

\[
(((\text{[WH } \alpha, ] \text{ [FIN do] he say) that}) \text{ [FIN, see]) it})
\]

a problem, though, since verbs are not assumed to license a saturated FIN element.

268
219) i. Santa took the magazine without paying for it.
   ii. What did Santa take without paying for?

Here I will explore very briefly how examples such as (219ii) might be accommodated relatively simply within the LG-based account of extraction described so far. From what I have already said it should be clear that processing of the matrix clause will proceed normally, yielding the structure in (220) where the verb “take” has been identified as a source of VT to the WH element of “what”:

```
(\{[WH α]\} [FIN do]) Santa take,)
```

Naturally, the parsing process will continue here until the adjunct clause, headed by “without”, has been linked to the matrix clause:

```
(\{\{[WH α]\} [FIN do]\} Santa) take,)
```

Once it has been established that there is no further word in P2 it will be clear that the verb “pay for” also has a spare licensing capacity. Under normal circumstances this would, of course, result in ungrammaticality:

222) *Santa took the magazine without paying for .

However in the case of (221), unlike (222), the spare licensing capacity of “pay for” is discovered within the context of a VT structure already established between “take” and the WH element of “what”. I thus assume that in (221) “pay for” can also be identified as a source of VT to WH. In this way the transferred dependency relation linking WH and α will actually have two separate sources.

223) `\{(\{[WH α]\} [FIN do]\} Santa) take,)
```

The discovery of the second source within the context of the first explains, I think, the immunity of ‘PGs’ like this to the effects of the adjunct island, first noted in section 1.3.4 of
the first chapter; throughout the processing of the entire adjunct clause in (223) the verb “take” in the main clause, which has no ‘true’ dependent of its own, is constantly available as a source of VT to WH. There will consequently be no processing hiatus while the second source is found. This of course contrasts with the situation in adjunct island violations, such as (119ii) and (129ii) above, where no VT-source is available in the matrix clause during the processing of the adjunct clause.31

There is no reason in principle why a VT structure should not have more than one source. More particularly, there is no reason why once one VT-source has been found any other word with a spare licensing capacity should not also be identified as a source for the same VT. In this way we could allow the creation of a VT structure to act as a sort of polarity-trigger for a sentence whereby any number of sources can subsequently be identified. This is of course a very free and unconstrained approach to multiple gapping, though one which is not without certain advantages. For example, contrary to what is often assumed, it is possible to find constructions which contain more than one PG (Ross 1967):

224) i. Santa hid the magazine after he took it without paying for it.

ii. What did Santa hide _ after he took _ without paying for _ ?

Although (224ii) is cumbersome, it does not strike me as being ungrammatical. In addition, some examples display two ‘gaps’ which bear different semantic roles:

225) i. Wally hoped Nyman would leave the hall before anybody could lynch him

ii. Who did Wally hope _ would leave the hall before anyone could lynch _ ?

Here then the α element of “who” will bear both the Subject role of “leave” and the Object role of “lynch”.

In my opinion examples such as (224ii) and (225ii) point to a less constrained and ‘fussy’ approach to multiple gapping than is usually found in other theories (Chomsky 1982,

31 Notice too that this account also successfully predicts the fact, again noted in section 1.3.4, that extracted adjuncts, unlike extracted arguments, do not license ‘PGs’:

A) **Why did Santa take a magazine _ [without paying for it _ ]?

In (223) the extracted argument, the α element of “what”, has two VT sources, and will thus be assigned two separate semantic roles, the Object role of “take” and the Object role of “paying for”. However, bearing in mind the discussion of adjunct extraction in section 5.6.1, if an extracted adjunct such as the α element of “why” were to have two VT-sources, as in (A), then this adjunct would have to assign a separate Subject role to each of the two words which d-commanded it. Given, then, that adjuncts, like verbs, will only ever assign a single Subject role, it is thus possible to explain the ungrammaticality of examples like (A), at least under the intended interpretation.
1986-a, Engdahl 1983, Pollard and Sag 1994 ch. 4). Naturally, there will be constraints on the number and nature of multiple gaps (or sources), but many of these constraints would appear to have more to do with matters of stylistics, pragmatics and even memory than syntax or parsing. Of course, what I have suggested here is not in any sense a fully developed account of multiple gapping, and there is undoubtedly much more to be said about this phenomenon. Nevertheless, I hope here to have at least offered an initial insight into the possibility that ‘PGs’ might be a relatively natural consequence of the application of VT.

5.7 - Conclusions

In this chapter I hope to have shown how the theory of LG may provide the basis of an effective and plausible account of extraction. As far as I am aware, this is the first specific examination of a full range of English extraction data to be conducted within a dependency framework. Moreover, I suspect that this may also be the first serious attempt to account for a body of data by reference to an integrated theory of dependency syntax and parsing.

In section 5.3 I outlined a relatively simple parser for LG, which I referred to as the LGP. This was shown to differ from other dependency parsers, notably in its pair-based central processing heuristic and its use of derived constituent structure to encode the order in which syntactic relations are processed. In section 5.3.4 I suggested that the pair-based operation of the LGP could derive the adjacency constraint, which is generally stated as a separate principle of the grammar in other dependency theories.

Section 5.4.1 introduced the notion of Valency Transfer (VT), the basic syntactic mechanism underpinning LG’s account of extraction. I suggested that VT, which allows the licensing capacity of one word to be transferred to another word in the same structure, could be stated simply and generally, theoretically allowing its free and unconstrained application, very much like the generalised transformation of Move-α in PP Theory. Naturally the application of VT has to be constrained in some way so as to prevent it from causing chronic over-generation in the grammar. In section 5.4.2, though, I raised the possibility that unlike Move-α the requisite constraints on VT need not be stated as syntactic principles, but can instead be shown to fall out independently from the operation of the LGP. I suggested that the parser might constrain the application of VT in two separate ways, which I referred to informally as ‘motivation’ and ‘locality’.
The motivation constraint, which arises from the parser’s natural inclination to build the simplest possible structure, guarantees that VT will only occur where it has to. In other words, the LGP will only go to the trouble of processing a VT structure if it represents the sole grammatical option. In sections 5.4.3 and 5.4.5 I argued that the motivation constraint restricts the application of VT to just two circumstances in English: i - to avoid violations of the head parameter by allowing FIN to license an extra pre-dependent, as in cases of topic extraction, and ii - to maintain the internal structural integrity of WH-words by allowing the WH element to continue licensing the α element with which it is fused. This applies in cases of WH-extraction.

The locality constraint on VT follows from the parser’s need to ‘see’ both the source and recipient of a transferred valency at the time of initiating a VT structure. In other words, the parser will not start processing a structure involving VT unless a word with a ‘spare’ licensing capacity is already available within the structure processed so far. Section 5.4.3 introduced the priority principle, another fairly natural axiom, which only allows a word to transfer its valency if it has no ‘true’ dependent of its own. I showed how this principle could interact with the locality constraint to ensure that in cases of long-distance extraction a series of consecutive VT relations was formed, the effect of which was to pass down the VT source locally from head to dependent through the syntactic structure as it was created. The resulting ‘hopping’ analysis, although reminiscent of WG’s visitor relation, was crucially different in that it only reflected the processing history of a structure rather than its actual syntactic form. In this way I argued that it is possible achieve the diachronic localisation of the long-distance relations arising from extraction without having to formulate any complex syntactic rules.

Perhaps the most interesting, and controversial implication of allowing the parser to govern the application of VT is that constraints on extraction will, inevitably, have to be explained in terms of processing difficulty rather than ungrammaticality. I examined this possibility in section 5.5 where I demonstrated that island effects could indeed be shown to result from the LGP’s inability to maintain a local relation between a VT source and recipient. The more stages in processing at which a local source was unavailable, leaving the extractee ‘stranded’, the more severe the island effect associated with that structure. I argued that this implied a scalar rather than absolute definition of ‘islandhood’ which accords very well with the attested variation in ungrammaticality associated with the data. I also suggested in section 5.5.4 that reformulating island constraints in this way may also allow for the possibility that
pragmatic information could intervene in the parsing process, serving to alleviate the effects of some island violations.

In section 5.6.1 I described how extraction asymmetries between objects and adjuncts could be shown to follow from the parser's response to their differing syntactic and semantic properties. In 5.6.2 I suggested that similar asymmetries between objects and subjects might in fact arise as a by-product of the specific syntactic property of complementisers in selecting saturated FIN as a dependent.

It may, of course, seem strange that what purports to be a defence of dependency syntax has relied so heavily on parsing considerations. However, as I mentioned in section 5.2.1, I believe that it is only through incorporating processing-related factors that an appropriately flexible account can be offered for island effects. Moreover, the important role played by the parser should not detract from the significance of the dependency syntax upon which it is based. After all, the core mechanism underpinning extraction is VT, a syntactic operation which is closely bound up with a dependency formalism. Indeed, I suspect that it would be virtually impossible to express any equivalent of VT in a phrase structure theory; VT involves altering the licensing capacity of one word by transferring the valency of another word. Given, though, that in constituency-based theories valency is registered directly in phrase structure patterns, any analogous process of VT would result in radical structural alterations in the environment of the source and recipient. For example, the recipient of VT in a phrase structure theory would require an additional specifier or complement position. Recall from section 2.3.2 of the second chapter, though, that it was precisely the need to avoid this sort of structural alteration that led to the development of trace theory and chains in PP Theory.

It is also important to bear in mind that many aspects of the account of extraction outlined in this chapter have also relied crucially on specific facets of LG syntax which were discussed in the fourth chapter. These have included assumptions concerning the internal structure of two-element words, differences between active and passive licensing, the revised head parameter as well as the system of derived semantic Subject and Object roles. Moreover, I believe that the LGP parsing system discussed in section 5.3 is really only compatible with a dependency grammar. Quite simply a phrase structure parser which operated on the basis of constituents would be unable to employ the basic system of derived constituency implied by cell structure to encode the processing history of a structure. In section 5.4.2 I described how this system of cell structure also plays an important part in the LG approach to extraction.
In terms of data coverage, I believe that the treatment of extraction outlined in this chapter surpasses WG and compares favourably with PP Theory. Indeed, the LG-based approach can successfully account for certain data which PP Theory cannot. For example, through the (stronger) motivation constraint discussed in section 5.4.2 it is possible to predict the uniquely leftward nature of extraction. Recall from section 2.3.2 of the second chapter, though, that PP Theory must stipulate the absence of rightward extraction in head-final languages where the c-command condition on chains would predict that extractees follow their traces. Moreover, the LG-based approach can account for all the island data discussed in the first chapter, including the adjunct and WH islands, for which WG offers no explanation, and the co-ordinate structure island which has been largely ignored in PP Theory. Unlike WG, the LG-based approach is also able to account for the full range of extraction asymmetries between subjects/adjuncts and objects. Here too, though, I believe that it can claim at least one advantage over the PP Theory account, since unlike the latter, it successfully predicts the absence of asymmetries between subjects and objects when the subject itself does not immediately follow a complementiser (see sections 1.3.3, 2.5 and 5.6.2).

Where I think that LG-based approach can be judged to be substantially better than both PP Theory and WG is in its flexibility; as I have already pointed out, a purely syntactic approach to extraction which implies an absolute binary distinction between grammaticality and ungrammaticality is simply inappropriate for the scalar nature of island effects. The more graded approach implied by processing difficulty is, I think, far better-suited to capturing these data. Moreover, I believe that the processing-related aspect of LG's treatment of extraction offers a more plausible interface with pragmatic criteria than a purely syntactic approach. This in turn holds out the hope that it might at some stage be possible to devise a fully holistic and integrated account of extraction which is able to accommodate some of the poorly-understood pragmatic effects first discussed in section 1.2.5 of the first chapter.

Naturally much work remains to be done in refining some of the ideas outlined in this chapter. For example, ideally it would be possible to explain the somewhat stipulative licensing properties of WH words discussed in 5.4.5 in terms of some more fundamental characteristic. So too the role of pragmatic information in guiding the parsing process needs to be explored further and formalised. More work also needs to be done on parasitic gaps. Nevertheless I hope to have fulfilled the central goal of this thesis by showing that a monostratal theory of dependency syntax can, after all, serve as the basis of a plausible and effective account of
extraction. I believe that this account goes much further than anything else currently on offer within the field of dependency grammar, and is thus in a better position to be judged alongside the PP Theory treatment of extraction data.
References


278


Ninio, A. (1993-b). 'Predicting the Order of Acquisition of Three-Word Sentences by the Complexity of their Dependency Structure'. Ms. The Hebrew University, Jerusalem.


284


286


