

DISCLAIMER OF QUALITY

Due to the condition of the original material, there are unavoidable flaws in this reproduction. We have made every effort possible to provide you with the best copy available. If you are dissatisfied with this product and find it unusable, please contact Document Services as soon as possible.

Thank you.

Due to the poor quality of the original document, there is some spotting or background shading in this document.

PAGES 105-107, 204, 206, 209 AND 218
ARE MISSING FROM ARCHIVES (AUTHOR'S)
COPY.

PARTS OF THIS THESIS ARE
ILLEGIBLE

LOCALITY PRINCIPLES IN SYNTAX AND IN PARSING

by

Amy Sara Weinberg

B.A.(joint first class honours) McGill University

(1976)

Submitted in the Department of
Linguistics and Philosophy
in Partial Fulfillment of the
Requirements of the
Degree of

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Signature of Author

Department of Linguistics and Philosophy
February 8, 1988

Certified by

Thesis Supervisor

Accepted by

Department Chairman

© Amy Weinberg, 1988

HUM.
MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

JUN 8 1988

LIBRARIES

VOL. 1

HUM.

LOCALITY PRINCIPLES IN SYNTAX AND IN PARSING

by

AMY SARA WEINBERG

Submitted to the Department of Linguistics and Philosophy on February 5, 1988 in partial fulfillment of the Requirements for the Degree of Doctor of Philosophy in Linguistics.

ABSTRACT

This thesis has three main aims. The first is to present criteria that can be used to constrain the class of natural language comprehension devices. We argue that a certain class of parsers; LR(k) bounded context processors can explain why grammars for natural languages must contain a Subjacency constraint. More generally, we argue that the ability to provide a functional explanation for linguistic constraints is a major criterion for judging the adequacy of classes of processing devices. In service of the first goal, we also consider several psycholinguistic experiments that have been taken as incompatible with the LR(k) approach and show that a more refined LR(k) theory (the Minimal Commitment Theory) can deal with these cases in an illuminating way.

Our second aim is to show that the theory of parsing forms an integral part of grammatical theory not only in the sense of providing a functional motivation for one of the theory's major constraints, but also in its ability to explain otherwise mysterious properties of subjacency. In particular, I claim that a parsing theoretic approach can explain why subjacency applies at S-structure, why it applies to movement, parasitic gaps, and a subset of gapping structures even though these structures do not constitute a natural class given only formal or linguistic substantive criteria. Thus the theory of parsing becomes an integral part of grammatical theory in that it, just like the functional demand of language learning dictates the form that grammars of natural languages can take.

The final aim of this theory is to propose an adequate formal theory of Subjacency and the Empty Category Principle (ECP). Following WAHL(forthcoming), I propose to separate the ECP into two parts: a condition of lexical government and generalised binding. These conditions apply conjunctively and in separate parts of the grammar. I argue for this approach by considering superiority effects and other "WH in situ" constructions. These conditions interact with the proposed subjacency constraint in that subjacency forces overt syntactic movement to be local. Local movement is allowed only if its output can satisfy both parts of the ECP. We argue that this approach yields a revealing treatment of the standard syntactic island effects, the Condition on Extraction Domains, movement from MPs, and the parasitic gap construction.

Thesis Supervisor: Dr. Noam Chomsky

Title: Institute Professor

To my Grandmother

Jeanne Kovner

who taught me that some things are worth doing no matter how long it takes. My love and thanks.

ACKNOWLEDGEMENTS

A friend of mine once suggested that I try not to be the first woman ever to receive a Festschrift before receiving her PhD. Though there are some days when I feel old enough at least for retirement, if not for the Festschrift, I think this document counts as taking his advice.

Perhaps the sole advantage of working on something for so long is that you have time to get help from an enormous number of people. Now it is time to thank at least some of them.

First I would like to thank Noam Chomsky for teaching me to be a linguist, for being an example of intelligence and integrity, for providing me with a framework in which to work, and for patiently waiting for me to finish.

Bob Berwick has been a sometime collaborator, thesis committee member, and fulltime friend. We began thinking about parsing and linguistic theory together and many of the ideas in this thesis stem from that early work.

Richard Larson agreed on very little notice to help supervise this work. He kept me honest in matters semantic and otherwise and suggested many helpful improvements.

David Lightfoot and Joseph Aoun also contributed mightily to this thesis. As half of "the WAHL framework", they helped develop the theoretical support for the analysis of Bounding that forms a large chunk of this work. Moreover they provided needed support and good ideas on demand. David also provided the collegial atmosphere that made it worthwhile for me to finish by starting the linguistics department at the University of Maryland.

Maria- Luisa Zubizarretta and Jean-Roger Vergnaud provided crucial insights and helpful comments particularly with regard to chapters Two and Three.

Janet Fodor deserves a big thank-you for forceful arguments that the deterministic theory of parsing was wrong. Large sections of this thesis are a response to her lucid criticisms. Paul Gorrell also provided needed scepticism and an alternative theory that was too good to be ignored.

I taught various pieces of this thesis at the Salzburg Summer Institute, the Universite de Quebec a Montreal, and at the University of Maryland. I would like to thank the students at all of these institutions for helpful comments and G. Drachman and the staff of the linguistics department at UQUAM for providing me with these opportunities.

Friends and family have kept me sane throughout this process. Thanks to Sari Hornstein, Joanne Kesten, Peter Murrell, Anne Riley, Shihab Shamma, Kim Sterelny and James Weinberg for good cheer and encouragement.

My parents deserve a big thank you for not asking (too often) about when I was planning to finish and for their love. Michel and Renata Hornstein provided much needed support and forbearance.

Finally, I would like to thank my sometime collaborator and fulltime husband, Norbert Hornstein for all of the stimulating discussions, for bearing with me through this, and for "being a truly remarkable person along every parameter worth considering."

TABLE OF CONTENTS

	Page
I. Introduction	7
II. Constraints on the Class of Natural Language Processors: Experimental Results, Determinism, and Bounded Context Parsing	13
2.1 Classes of Processing Devices	14
2.2 On the Simulation of Perceived Processing Complexity	17
2.3 Empty Operators and Non-Parallel Processors	23
2.4 Left Context and the Analysis of Operator Variable Structure	29
2.5 The Minimal Commitment Theory	36
2.6 An Apparent Problem for Non- Backtracking Models	45
2.7 Some Advantages of the Minimal Commitment Theory over Backtracking and Parallel Devices	48
2.8 Conclusion	60
Footnotes to Chapter 2	62
III. Subjacency and the Deterministic Model	68
3.1 Introduction	68
3.2 Bounded Context Parsing	71
3.3 The Problem of Ambiguity and Left Context for Bounded Context Parsers	77
3.4 Parasitic Gaps	84
3.5 Licensing of Parasitic Gaps	90
3.6 On the Distribution of Epithets	92
3.7 Critical Remarks	96
3.8 Gapping Constructions	100
3.8.1 Description of the Problem	100
3.8.2 Ambiguity and Gapping	106
3.8.3 Complement and Adjunct Attachment	112
3.8.4 Subjacency as an S-structure Condition	114
3.8.5 Subjacency as a Natural Condition on Grammars	118
3.9 Bounding Conditions in LF	119
3.10 Conclusion	129
Footnotes to Chapter 3	130
IV. Two Versions of Locality	138

	Page
4.1 The <u>Barriers</u> Theory of Government and Proper Government	139
4.2 The WAHL Framework	151
4.3 Conclusion	175
Footnotes to Chapter 4	177
V. Two Notions of Bounding	189
5.1 Bounding in the <u>Barriers</u> Framework	189
5.2 Bounding in the WAHL Framework	196
5.3 Concluding Remarks	212
Footnotes to Chapter 5	215
VI. A Comparison of the <u>Barriers</u> and WAHL Frameworks	219
6.1 Superiority	220
6.2 Pronouns as Bound Variables Revisited	229
6.3 General Theoretical Consequences of the Account	
6.3.1 Movement from NPs	235
6.3.2 Movement from NPs in Romance	243
6.4 Conclusion	250
Footnotes to Chapter 6	252
VII. On the Distribution of Parasitic Gaps	267
7.1 Introduction	267
7.2 Indexation within the Wahl Framework	269
7.3 Subjacency and Parasitic Gaps	270
7.4 The <u>Barriers</u> Approach to Parasitic Gaps	277
7.5 Anti-C-command and Parasitic Gap Licensing	281
7.6 S-Structure Licensing	285
7.7 A Surprising Assymetry	287
7.8 Other Approaches	288
7.9 Conclusion	292
Footnotes to Chapter 7	294
Bibliography	296
Biography	307

Introduction

This thesis investigates topics in the theory of grammar and the theory of parsing. Typically, accounts in the psycholinguistic and computational literature provide brief analyses of linguistic structures and propose accounts that are not heavily dependent on what linguists know about the analysis of natural language. In this thesis, we take the opposite tack, and ask what the consequences would be for a parsing theory if it took the analyses proposed by linguists seriously and tried to incorporate them into parsing analyses. We also look at the question from the other direction, asking what linguistic theory can learn from consideration of parsing models.

Linguistic theory has taken the solution to the so-called "Logical Problem of Acquisition" as its research goal. That is, children are exposed to data that is both deficient and degenerate; not containing relevant data to support the inductive generalisations that children eventually arrive at, and containing poorly presented exemplars of evidence. Nonetheless children attain a rich system of knowledge in a surprisingly short time. Linguists have thus taken their task to be the provision of an innate linguistic faculty that can guide the child through his impoverished environment to

the adult state of knowledge about language in a relatively short period of time. Solving this problem places heavy empirical constraints on the form that the innate linguistic endowment can take.

We do more with our language however, than learn it. We are also able to produce a semantic representation for a sentence in more or less the time that it takes us simply to hear (or read) the words of the sentence. Considering the complexity of the mapping between the phonetic string and the semantic interpretation, this is a remarkable feat. One might wonder whether the solution to this problem would also place constraints on the form of the innate endowment. This thesis explores this possibility; arguing that the presence of the Subjacency Constraint proposed in Chomsky(1973, 1981, 1986) in Universal Grammar is to be explained in terms of the role it plays in allowing the natural language parser to build the appropriate syntactic analysis to support semantic interpretation in an efficient way.

Looking at things in this way also allows us to decide which, among a wide range of possible processing devices models the human language comprehension system. This is an important side benefit because, as was the case in generative grammar several years ago, many mutually incompatible parsing approaches seem fully capable of describing a wide range of

experimental data. Thus we need evidence of a different sort to tell us which device is the appropriate model for human beings.

The second part of this thesis looks at the interaction between the parsing theoretic motivation that we provide for the subadjacency constraint and the actual formal analysis of subadjacency in a variety of grammatical approaches. Some of these analyses are compatible with our interpretation of this condition while others are not. We present the various analyses and try to argue on linguistic grounds, for an analysis that is compatible with our interpretation of this constraint.

Let me now present a brief tour of the chapters of this thesis:

Chapter Two presents a variety of mutually incompatible parsing models for natural language. We present a range of experimental data for which all models are descriptively adequate. Further inspection reveals that only one model: a deterministic parser employing a minimal commitment strategy is fully compatible with a wider range of experimental data. We present the organisation of this type of parser and motivation for its design properties.

Chapter Three presents a formalisation of a deterministic parser in terms of an LR(k) parser; a model developed in Knuth(1975). We show that a version of this model, a bounded context LR(k) parser would require a locality condition like subjacency in order to guarantee efficient parsability of WH moved structures. We argue that looking at subjacency as a locality constraint motivated by parsing considerations also allows us to explain the set of constructions that this constraint applies to: WH and NP movement, parasitic gaps, and gapping constructions. The constraint does not apply to LF movement, binding, or control structures. Looked at from the perspective of the grammar, the class of constructions that subjacency applies to is unnatural. We will see however that all of these constructions confront a bounded context parser with the same problem; a problem that the parser can solve by invoking a subjacency restriction.

Chapter Four presents the grammatical framework in which current formal analyses of the subjacency constraint are set. We review the basic idea proposed in Barriers (Chomsky(1986)), which is the unification of the domains that are relevant for the theories of government, proper government and bounding. We propose an alternative system that unifies antecedent government with the binding theory: an idea proposed in Aoun (1985) and developed in Wahl (1987). Under this approach we would not expect the same

domains to be relevant to "antecedent government" and bounding theory.

Chapter Five explores the bounding theories that result from the assumptions of the Barriers and WAHL frameworks. We try to show that the unification achieved in Barriers comes at the expense of making a series of unmotivated assumptions. We pursue the idea of unification proposed by Chomsky, but within the WAHL framework. We show that in the Wahl framework, there is an intimate connection between the theory of lexical government and bounding for the case of overt movement. We derive the traditional syntactic island phenomena first noticed in Ross (1967) in both frameworks. We also show how the Condition on Extraction Domains from Huang(1981) would fall out from the subjacency theories proposed by both of these systems.

Chapter Six explores some differences between the Barriers and WAHL frameworks in an effort to choose between them on empirical grounds. The main difference between these frameworks is in the treatment of antecedent government. Antecedent government is taken to be a subspecies of government in the Barriers system and a subspecies of Binding in the WAHL system. We examine a raft of LF phenomena, presenting an analysis of so-called superiority effects and reconstruction cases. We show that many problems associated

with these constructions disappear if we treat the LF representations for these cases as structures of unrestricted quantification. We show that this idea is compatible with the WAHL approach but is not compatible with the definition of antecedent government proposed in Barriers. We also present another class of problems for this approach by looking at movement out of subject noun phrases in Romance languages.

Chapter Seven deals with the parasitic gap construction. We show that the definition of subjacency proposed in Barriers to handle these constructions is incompatible with our view of this condition as a locality constraint imposed by considerations of efficient parsing. We propose an alternative consistent with our approach, present some empirical support for our theory, and compare it to other theories in the literature.

Constraints on the Class of Natural Language Processors:
Experimental Results, Determinism and Bounded Context Parsing

One of the major problems facing the theory of natural language processing is the fact that a wide array of experimental paradigms from the psycholinguistic literature seem consistent with a host of mutually incompatible parsing algorithms and even contradictory parsing architectures. Thus, while data from psycholinguistics can tell a lot about the way that people process natural language, such data do not have, enough constraining power by themselves to distinguish the unique architecture and class of algorithms that underlies the human natural language processing device from the general class of possible architectures and algorithms. I will first discuss a general classification of the basic types of processing architectures. Next I will relate these models to psycholinguistic data by considering the interpretation of questions, showing first what experimental data tell us about how human beings comprehend questions, then discussing the class of architectures and algorithms compatible with this data. Next, I will discuss a special class of parsing mechanisms, bounded context deterministic parsers, and show how they can be made compatible with theories of linguistic representations by having them process according to a "minimal commitment" scheme. Finally, I will discuss a wider range of

psycholinguistic evidence that seems to favor this "minimal commitment" approach. Additional assumptions about the parsing mechanism are also introduced and psycholinguistically justified.

2.1 Classes Of Processing Devices

The basic classes of natural language processors are defined with respect to the way that they handle local ambiguities in natural languages. Local ambiguities occur in cases like (1):

(1a) Mary expected Fred.

(1b) Mary expected Fred to arrive.

The sentence is locally ambiguous because at the point when the device is analyzing the lexical item 'Fred', it cannot tell whether this item is to be attached as the direct object of the verb 'believe' or as the subject of the complement clause 'Fred to arrive'. The structures are given in (2a) and (2b).

(2a) [Mary [_{VP} expected Fred]]

(b) [Mary [_{VP} expected [_S Fred to arrive]]]

This sentence cannot be disambiguated until the complement verb is detected at a later point in the parse.

Two classes of parsers deal with this problem by adopting a "he who hesitates, is lost" strategy. Upon meeting an ambiguous fragment of the sentence a parallel parser pursues all possible analyses consistent with the material that it has previously encountered. In a case like (1a), a parallel parser would initially construct both the analyses 2a and 2b because each is consistent with the material seen until it reaches the predicate 'to arrive'. Once such a parser reaches a disambiguating portion of the sentence, it drops the analysis that is inconsistent with the disambiguating material.

A second type of parser is called a backtracking parser. When confronted with an ambiguity like (1) above, this parser pursues one analysis based on some predetermined stored protocol. That is, the parser can be preprogrammed to choose to pursue i.e. an analysis like (2a) as its first choice for any sentence containing a local ambiguity as in (1). It does so without consulting the actual input to see whether the choice is in fact the right analysis for the case at hand. In the case discussed above, the parser will of course be wrong if the sentence turns out to end as in 2b. At this point, the parser backtracks from the disambiguation point,

to the point where it made its initial misanalysis and reparses in accordance with the data that it currently has available.

A third type of parser takes the more conservative "look before you leap" approach. This type is called a deterministic parser. Once a deterministic parser proposes an analysis, it is stuck with it in the sense that it can only pursue one analysis at a time, and once an analysis is postulated, it cannot be rescinded when later disconfirming evidence is encountered in a parse. Marcus (1980) proposed such a parser based on the Extended Standard Theory and proposed two conditions that defined it as deterministic.

"First, all syntactic substructures created by the grammar interpreter are permanent. This eliminates the possibility of non-determinism by backtracking... Second, all syntactic substructures created by the grammar interpreter for a given input must be output as part of the syntactic structure assigned to that input ...this property eliminates the possibility of simulating non-determinism by pseudo parallelism."

2.2 On the Simulation of Perceived Processing Complexity:

Because of sentences like (1) and a variety of ambiguous sentences like those listed in (3), Marcus claimed that imposing a determinism condition forced the addition of "a lookahead" mechanism to the parsing architecture. A lookahead device allows the processor to delay some parsing action that it is not sure how to compute correctly while it scans forward to gather information that will tell it exactly what the next correct move should be.

- (3) a. Have the students take the exam!
- b. Have the students taken the exam?
- c. The man [up] ran after the girl]
- d. The man up[ran] [p]after his dinner]

The sentences in (3) are all ambiguous at various points, and each sentence is only disambiguated at some later point in the analysis. In addition, all of the analyses of the sentences in (3) seem to be quite easy to process; a fact that the processing model must be able to simulate. This rules out an analysis where a deterministic parser would pursue the wrong analysis, reach a dead end, and have to start over from the beginning, because we don't have the conscious sensation of an actual parsing failure for these cases. Examples like (3) contrast with cases like (4) where

there is a conscious sensation of difficulty and in some cases, even failure to come up with the appropriate reading.

(4) The horse raced past the barn fell.

Marcus (1980) explained the differences between the two types of cases by bounding the size of the lookahead. He correctly observed that the lookahead had to be bounded in order for the determinism hypothesis to have any content.

A bound on access to upcoming input (right context) is necessary to make a processing device psychologically plausible. Miller (1951) showed that short term memory could only hold a minimal amount of uninterpreted material. This is what lexical items that are scanned by the lookahead and that have not yet been associated with a logical role by being incorporated into a syntactic tree are. In addition to the conceptual and general considerations about the structure of memory, Marcus (1980) further claimed that the structure of the lookahead device could also account for the differences between sentences in (1)-(3) and (4). The basic structure of the Marcus parser is as follows: Lexical items are accessed by passing through an input buffer. This buffer can access between 3 to 5 words at a time.¹ The parser analyzes incoming material by comparing it to a set of stored pattern-action rules. These rules can access any element that the parser can store in the input buffer and the parser

can look within the buffer space to see which rule pattern matches the input lexical items. The rules tell the parser how to construct a syntactic tree representation. The parser can also access a bounded amount of material that it has previously seen (left context). Marcus also limits the amount of left context material to those elements contained in the same sentence as this constituent.² This type of architecture as well as the backtracking architecture divides cases into three types. It seems that some bound should be placed on left context because we cannot remember infinite stretches of previously encountered structure. Let us now review how the Marcus parser deals with different types of structures.

(a) Cases like (1) and (3) that are locally ambiguous but cause the parser no difficulty.

In these cases backtracking models claim that the backtracking needed to transform an incorrect false start into a correct analysis is so minor that it is not associated with a computational cost. The determinist claims that these cases involve such a minimal lookahead that disambiguating information is sure to be in the lookahead buffer and so the parser will not make a mistake on this input. A mixed solution would supplement a backtracking parser with a minimal lookahead. These parsers use an exact analogue of a

deterministic parser's local buffer solution and thus always make the right choice in these cases. Such a solution would work for cases like (1) and (3b) (3c). In a case like (1b) even if the parser mistakenly hypothesized that the subject of the embedded infinitival was the direct object of the verb 'believe', the backtracking needed to insert the infinitival S marker between it and the verb is minor and a nondeterministic parser might correct its mistake in a way that came "cost free". By analogy, the lookahead needed to figure out whether the matrix had a sentential or nominal complement is minimal, as the point of disambiguation is only one lexical item away from the point of initial ambiguity. Thus on the deterministic story, all interpretations for these sentences should be easy to analyse.

(b) In contrast, there are cases that require more extensive backtracking over essentially unbounded distances. These cases can be divided into two types.

There are those for which people register a strong preference for one of the possible analyses (even when pragmatic biasing points to the other choice but where both readings are eventually made available.)

An example of this case is shown in (5), where, as Fodor 1985 mentions, there is an initial preference for the

reading where 'who' is taken to be the object of the embedded preposition.

(5) a. Who_i did the little girl beg e_i [_S PRO_i to sing
those stupid French Songs?

b. Who_i did the little girl beg to sing those stupid
French songs((for) e_i)?

The misanalysis (what is referred to as a "garden path") occurs in (5a) because a trace is not placed in the post matrix object position. Since the sentence ends without another position from which to interpret the Wh phrase, it is initially taken to be unacceptable, subject to reanalysis. The sensation of conscious difficulty results from the fact that the point of disambiguation (the final punctuation point) is quite far away from the point where the parser must decide whether to insert a trace or a PRO into the subject position of the embedded sentence. Within a backtracking parser, this means that the device must backtrack over a significant distance to change its initial insertion of a trace in embedded subject position to a PRO. The deterministic framework also equates the difficulty with (5b) to the distance between the point of ambiguity and the point of disambiguation. The disambiguation point is simply outside the scope of the lookahead buffer and so the parser

makes the wrong choice at the ambiguous point. Since parsing decisions cannot be rescinded, the device is stuck with this mistake, cannot find an object gap with which to co-interpret the WH-element, and goes into an error condition.

c. The other cases are the conscious garden paths where one reading is so difficult to process that it usually has to be explicitly pointed out to people, even if it is the only reading that results in a grammatical sentence. There are cases like (6).

(6) The horse raced past the barn fell.

The processing load associated with cases like (6) is compatible with the backtracking approach if it is assumed that backtracking over long distances is computationally costly. The extra burden imposed by true garden paths is a complex effect that is partly lexical, partly structural and exacerbated by distance effects.

We will discuss these cases and their relevance to the choice of the correct natural language processing device below.

The straightforward prediction of a parallel parser is that all possible interpretations associated with the above sentences are computed during the initial analysis phase. Therefore relative difficulties in retrieval of a particular

analysis must reflect the difficulty of communicating some particular analysis to a decision-making part of the system that analyses natural language. We will discuss a proposal of this type below.

2.3 Empty Operators and Non-Parallel Processors

The following sentences (first pointed out in Fodor (1985)) present a superficial problem to both deterministic and backtracking parsers. In factive constructions, and other constructions that are analysed by predicating an empty operator headed clause to a head, the presence of the head makes creating an empty operator possible but it does not make this position obligatory in these structures. Consider (7) and (8).

7. The fact that Mary stated clearly was surprising.

8. The fact that Mary loved Bill was surprising.

In a case like (7), the parser must place an empty operator in the complementiser of the relative clause in order for the relative clause to be interpreted as a predicate and so that the operator can bind the empty object of 'stated'. In (8) by contrast, we do not want to place an empty operator in either adjunct or relative clause initial position because

there is no empty position for the operator to bind. In (8) there is no corresponding gap position. Because of the possibility of successive cyclic movement however, the gap can be indefinitely far away on the surface from the empty operator position. A deterministic parser with limited lookahead will not be able to wait for the disambiguating right context. Therefore, there will be certain cases where it will incorrectly place an empty operator in the adjunct's COMP.

Fodor 1985 implies that these facts pose a problem solely for deterministic parsers, suggesting that a nondeterministic solution is called for. In fact the determinism/nondeterminism issue is beside the point. Both deterministic parsers and non-deterministic parsers with backtracking are going to have problems in these cases because they both at least superficially predict that such cases should cause people to have noticeable difficulties in comprehending these sorts of sentences. However, neither of the sentences in (7) or (8) cause any difficulty at all. Thus these sentences break the association inherent both in deterministic and backtracking architectures between the length of material intervening between an ambiguous point and the point of disambiguation in a sentence and perceived processing difficulty. I would like to suggest a solution for this

problem that is compatible with either one of these approaches.

We could solve these problems if we could design an algorithm in which the semantic component simply didn't interpret empty operators unless they were eventually bound to elements in argument positions. Since these elements have no phonetic content, if they received no semantic interpretation, it would be as if these elements never existed. In that case we could insert the empty operator in all cases but we would always be sure to be right because an unbound empty operator would simply be ignored because it was invisible. There are several ways to implement this suggestion. The first is a loose translation of visibility conditions discussed in the linguistic literature into conditions on the processing mechanism. Loosely following a suggestion by Aoun (1981), subsequently adopted by Chomsky (1981) we could claim that different interpretive components (LF and PF) can only detect material that is coded within the appropriate vocabulary for these components. Concretely, this means that the PF component can only interpret categories that bear the appropriate phonetic features (like Case) and the LF component can only interpret categories that have semantic features (categories that are referential, delimit a range or are associated with a predication or thematic role.)

In Berwick and Weinberg (1984), we provided an implementation of this idea in terms of a two stage parsing model. The first stage dealt with tree expansion and built an s-structure analysis of the sentence. The second stage dealt with interpretation of structures and could search the previously built syntactic representation. We can think of this 2nd stage representation as an analogue to the LF component of grammar i.e. as a component that provided a semantic interpretation for previously built syntactic structures. Pursuing the intuition that the 2nd level "LF" was a representation concerned with purely semantic aspects of the interpretation, we placed a semantic visibility condition on the categories appearing in this component. We claimed that, to be interpreted by the "semantic component", a category had to have semantic features. These were the features that allowed a noun phrase to either denote an individual or set of individuals or allowed a quantifier to delimit a range. Assuming a category had such features it would be given a "referential index" and be visible in the LF component. If a category did not intrinsically have such features, it could obtain a referential index by being linked to an element that did. Given the visibility condition, an element would have to receive a referential index before being shunted into the LF component. If an element did not receive an index before shunting, it would become invisible

and receive no interpretation. We will state the combined condition on visibility as follows:

VISIBILITY CONDITION

To be visible in the interpretive (LF component), an element must be associated with a theta role [either by occupying a theta position or binding an element in a theta position]. The output of the LF component must also associate such a category with an element that has referential features [features that either designate an individual or set of individuals or that delimit a range].

Given this idea, it is easy to see why cases like (7) or (8) do not cause any processing difficulty. We will go through these cases here simply to make the analysis absolutely concrete. The parser will create an empty operator in the initial position of the embedded clause in (7) or (8) because both structures could contain an empty position later on in the sentence. Thus we will derive structures like (9) for both examples.

(9) The fact [_S O_i that[...

In case (7), the structure will ultimately be linked to a trace in a thematic position as in (10). It will also get a referential index by being associated with the head of the

relative . Thus it satisfies both conjuncts of the Visibility Condition proposed above and so the structure, complete with empty operator is visible in the interpretive component.

(10) [_{NP} [The [_N fact [_S O_i that [_S Mary explained e_i clearly]]]].

In (8), all theta positions are filled by elements that are unlinked to the empty operator. This is shown by the structure given by (11). Since the empty operator does not bind a thematic position, it cannot be interpreted by the LF. It is as if the operator had never been postulated.

(11) [_{NP} [_{NP} The fact] [_S O_i that [Mary loved Bill]].

This type of analysis allows us to retain the connection between distance separating an ambiguous point in a parse from disambiguating material and complexity of comprehension of a structure that is implied by both backtracking and deterministic models and which seems to make the right empirical predictions for a large class of cases. In section 2.5 below, we shall present some further experimental facts that are inconsistent with parallel models and suggest that the empty operator algorithm is therefore a necessary part of a fully psycholinguistically plausible parser.

2.4 LEFT CONTEXT and the Analysis of Operator Variable

Structure:

The analysis of cases like (7) and (8) also implicitly assumes that the creation of a variable (or its trace) in an empty position is triggered by the presence or absence of an operator, either phonetically specified or empty in the sentence. That is, we only want to expand a phonetically empty position in a case like (12), if we have previously encountered a question word in the sentence; i.e. in (12a) but not (12b).

- (12) a. Which boy did you think that Mary believed that Fred
decided to run in the race?
- b. Did you think that Mary believed that Fred decided
to run in the race.

The problem here is similar to problem of creating empty operators but in the reverse direction. The WH or empty operator can be indefinitely far away from the potential variable. If a backtracking parser literally had to backtrack over all the material that could possibly intervene between the variable position and the operator, we would predict that questions like (12a) or (12b) should be very difficult to interpret. In fact, neither sentence is much

more difficult than an active sentence of the same length, and the extra difficulty seems to be no greater than that between a short question and its corresponding active. The same predictions apply within a deterministic framework, if the device contains a bounded, and thus psycholinguistically reasonable mechanism to search previously encountered left context, the WH operator will be beyond the scope of this mechanism in a case like (12). This will entail that it will insert a variable based simply on information that is locally available in the phrase structure representation, incorrectly predicting that one of the readings in (12) should be impossible or at least extremely difficult to comprehend.

It seems clear that either a backtracking or deterministic parser must be supplemented with a mechanism for locally encoding the presence of the WH operator so it can later be retrieved in a relatively cost-free way. In fact there are basically two mechanisms that have been proposed to do this: methods of generative and literal encoding.

We will first discuss two generative encoding mechanisms. The first was proposed by Wanner and Maratsos (1978) and is known as the "HOLD Hypothesis." Under the HOLD analysis, the parser places a copy of a WH-word into a special holding bin at the point when that WH word is first encountered in a sentence. The element remains in this bin until a potential

variable position is located later in the string. At this point, rather than backtracking over left context, or guessing about whether or not to insert a variable, the parser simply accesses the HOLD cell. Its actions are guided by the contents of this part of the mechanism. If the HOLD cell contains a WH word, then the parser can infer that a WH element was previously encountered in the analysis and insert a variable in the potential variable position of the phrase structure representation. If there is nothing in the HOLD cell, then no variable will be inserted in this position. Since the presence or absence of a previously encountered WH element is not retrieved from the previously built phrase structure representation, then the literal distance between the question word and the variable position should have no effect on the difficulty of interpreting a question.

A similar result can be achieved by generatively encoding the presence of a WH element through the phrase structure rules. This method involves annotating the phrase structure rules so that each rule encodes whether or not an operator has been previously seen in the sentence. Each rule of the grammar can be paired with a duplicate rule that simply indicates that this phrase structure rule is also a possible expansion site for a variable. This is indicated by annotating the PS rules with a slash notation. Categories that appear to the right of the slash are interpreted as

potential variables to be inserted into the phrase. A term that has the same category to the right and left of the Slash is interpreted as a variable. This method was first proposed in Harman (1963) and is currently the method employed in Generalised Phrase Structure Grammars. We will discuss this method in more detail in the next chapter.

A final method involves literally encoding the presence of this question word throughout the phrase structure. In this case, this comes down to adopting a parsing analogue of Chomsky (1973)'s theory of successive cyclic movement. Bresnan(1972) discussed the left (or right) peripheral categories (complementisers) that sentences contain and that serve as marks of subordination. Examples are given in (13).

- (13) a. I believe that Mary is a good scholar.
 b. I want very much for Mary to be elected.

Question words may also appear in this position as shown by (14).

- (14) I wonder who you think is a great scholar.

Combining these facts, Chomsky (1973) assumed that complementiser positions, like all other grammatical categories could be filled with empty copies of any lexical

element that could fill the position. In fact, given the precursor of the Subjacency restriction discussed in the next chapter, he assumed that the representation of (12a) would be (14b).

(14b) [_S Which boy_i do[you think [_S e_i that [Fred decided [_S e_i [PRO to run e_i in the race]]]]]

Recoding this as an algorithm, the parser's first action upon opening an embedded clause would be to check the complementiser of the previous clause to see whether it contained a WH element or a trace. If the parser found a WH element in COMP position, it would insert a trace into the current complementiser position. The process would iterate until the chain of traces was associated with a theta position in the syntactic string. To execute this analysis, the parser only has to have access to the previous clause at any point in the parse in order to interpret questions efficiently and correctly. Thus the search of left context is literally bounded.

One might think that psycholinguistic evidence would be able to tell us which of these algorithms is actually used by the human comprehension device. The problem is that, while psycholinguistic experiments clearly suggest that subjects do compute operator variable structures as they interpret

questions, all of the algorithms mentioned above can be made to comport with the experimental data. I will give two examples of the type of evidence that I have in mind. Wanner and Maratsos (1978) reasoned that in order for subjects to interpret questions as operator variable structures, they had to remember that they had seen an operator when they reached a potential variable position. This meant that they had to retain the uninterpreted WH element in memory from the time it was first encountered in a parse until the time it reached the variable position. Recall that Miller (1951) had independently established that the storage of uninterpreted elements placed a heavy burden on transient memory. The HOLD hypothesis thus predicted that subjects would be relatively worse at performing some task that accessed transient memory if they were probed during the period between the point of discovery of the WH element (and its placement as an uninterpreted category into the HOLD store) than if they were probed after encounter of the variable position, (interpretation of the element in HOLD by association with a theta position and removal of the element from HOLD). They tested this prediction by presenting their subjects with a series of relative clauses. Subjects were told that they would have to read these sentences for comprehension, as this would be tested during the experiment. At various random points during the presentation of these sentences, Wanner and Maratsos also presented their subjects with a list of

unrelated proper names that subjects were told that they would have to recall. The dependent variable was the number of names from the list that subjects could remember. Sometimes the list of names appeared after the operator was associated with the variable position and sometimes it appeared before. Examples are given in (15). The "#" symbol indicates points at which the list of unrelated names appears in the full set of stimuli. Each particular sentence was interrupted at only one of these points.

(15a) # the witch who_x despised # sorcerers frightened #
little children#

(15b) # the witch whom_x sorcerers# despised e_x frightened #
little children#"

As discussed above, the HOLD hypothesis predicts that subjects would be better at the memory task in the former case. This is the result that Wanner and Maratsos obtained. Notice though that all the algorithms cited so far make this prediction because they all involve holding an uninterpreted element in memory until a theta position is reached. Thus the Wanner & Maratsos experiments can't tell us which algorithm we really use.

Another experiment that makes this point is a rhyme priming experiment published in Tanenhaus, Carlson, and Seidenberg (1985). Tanenhaus et al found that subjects recognised a visually presented target word that rhymed with

a previously presented prime faster than if the context contained no rhyme. They demonstrated that this effect disappeared if at least seven words intervened between presentation of the rhyming word and the target.

Interestingly, rhyme priming remains effective beyond the seven word limit if these lexical items intervened between a WH operator and its variable position.¹⁵ As in the Wanner and Maratsos experiment, Tannenhaus et al take their results to tentatively suggest that "...listeners appear to hold on to the verbatim form of the filler word until the gap is identified and filled¹⁶

To make matters worse, any of the algorithms proposed above can be embedded in a parallel, backtracking, or serial architecture. Therefore the available experimental evidence seems to hopelessly underdetermine the underlying structure of the natural language processing device and the algorithms that this device uses.

2.5 The Minimal Commitment Theory

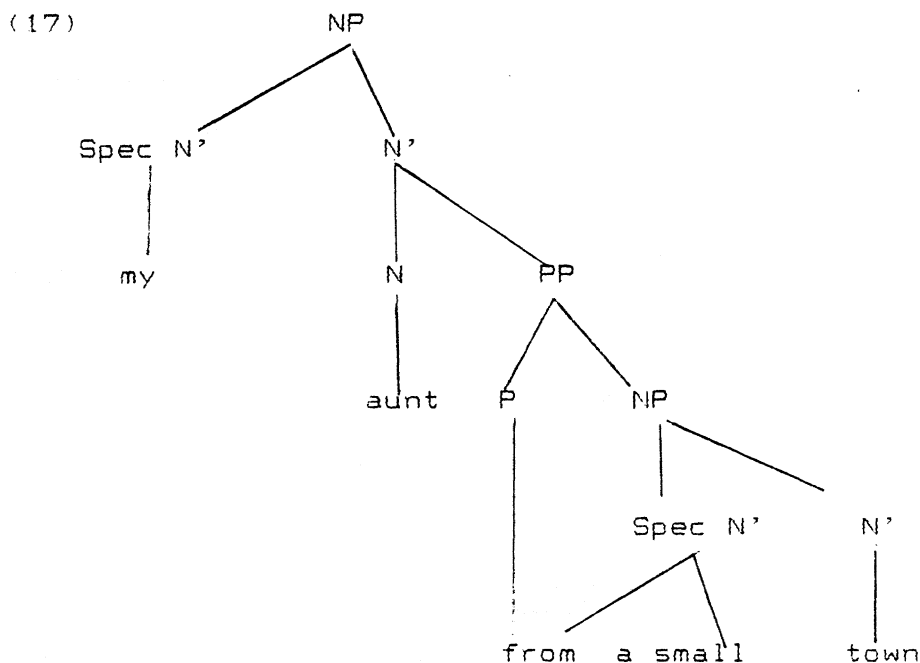
Before subjecting the deterministic, parallel and backtracking parsers to further experimental test I will look more closely, at the properties of a deterministic parser arguing that if the human natural language processing device

is a deterministic device, it is one that does not make use of a lookahead.

The first problem with the lookahead mechanism is that, given the structure of a Marcus type model, we must either make the lookahead buffer so capacious as to be psycholinguistically implausible, or incorrectly predict that certain sentences that are in fact quite easy to process will cause the processor to make mistakes. This is because disambiguating information appears outside of the scope of the lookahead window. (16) is a case of this type:

(16) I drove my aunt from a small town in Wisconsin's car
into the city.

After the parser has recognized a verb phrase by seeing the verb 'drove', the buffer would contain the lexical items spanning from 'my' to 'town'. Thus the parser can assume at most that it has seen a complex noun phrase with the following shape :



If we hypothesize the attachment of the complement right after it has recognized the verb, then it will incorrectly assume that (17) functions as the direct object of the verb 'drove'. Notice that we can expand the complement structure of this sentence indefinitely and so no matter how much we expand the parser's lookahead it should still incorrectly make mistakes in these cases. We can build parser routines to handle these phrases without making mistakes in a number of ways. First, we could wait until we come to the end of the NP 'my aunt from a small town in Wisconsin', and build this phrase before deciding where to attach it in the tree structure. This is the solution adopted in Marcus (1980). If we adopt this solution though, we commit the parser to

storing potentially large amounts of material whose logical role is unknown. Following Miller (op.cit.) this should quickly overburden short term memory and we would incorrectly predict that sentences with constituents of this type should be difficult to understand. We could also bound the domain of uninterpreted constituents that we allow the parser to store by forcing it to attach a structure to a tree within a bounded distance. The parser is then bound to err in cases like these because the evidence telling it that the NP should be attached as a determiner to the complement of 'drove' and not directly as the complement of this verb will be outside of the buffer's view at the point of attachment."

A variety of authors, (see Frazier and Rayner (1982) Fodor (1985), and Berwick and Weinberg (1985)) point out that a parser with even minimal lookahead should be able to process short garden path sentences like (18) easily. For example, in order to have enough lookahead to process the embedded clause in (18)^a the parser must be able to store at least 3 lexical items in memory. Unfortunately, this would give the parser enough access to process a sentence like (18b) correctly because the device should be able to detect the final verb of the sentence and thus realize that the first verb was part of an initial relative clause.

(18a) The man believes [the woman really is a spy].

(18b) [_S [_{NP} The boat floated] sank]

Marcus et al (1983) deal with cases like (16) by modifying the structures that the parser builds. Instead of building syntactic tree representations directly, they adopt a formalism that is reminiscent of the one proposed by Lasnik and Kupin (1977). This formalism allows one to capture all of the relations that one states using syntactic trees, without constructing these trees directly. The model represents syntactic information as a series of statements. Separate statements express linear precedence and dominance relations. More importantly, the only predicate used to express hierarchical relations is the predicate dominates. The notion of "direct domination" can be inferred from the representation. (If X and only X dominates Y, then one knows that X directly dominates Y.) The expression 'X dominates Y' is written as in

(19)a and 'X precedes Y' is expressed as in (19b).

(19)a. $D(X,Y)$

b. $X < Y$

Using this representation helps us to do away with the need for a lookahead buffer because it allows the parser to be conservative. The parser can say that it knows that one

category dominates another at an ambiguous point in a structure, even if it does not know whether or not there is a relation of 'direct dominance'. With this in mind, let us reconsider a case like (2), which we repeat below which Marcus claimed force a deterministic parser to include a lookahead device.

(20)a. Mary [_{VP1}expected [_{NP2}Fred]].

b. Mary expected Fred to leave.

The parser can use the information contained in the lexical entry of 'expect' to tell it that this verb must take a complement and it knows that 'Fred' is part of a noun phrase. Putting these two pieces of information together, it can predict that the NP will be dominated at some level by the VP containing 'expect', either as the verb's direct object or as the subject of the verb's complement sentence. Thus it can add (21) to its assertion set.

(21) D(VP1, NP2)

Mary < expected < Fred....

In a case like (20b), the parser will have to add the statement (22) to the assertion set after it detects the presence of the infinitive.

(22) D(VP1 S2)

D(S2 NP1)

Notice that adding this statement doesn't force the parser to retract any of its previous assertions and thus the parse for this sentence is deterministic."

Finally, as proposed by Berwick and Weinberg (1985), and Berwick (forthcoming) it is helpful to let the parser know when it has finished a phrase and can thus close a domination statement. To do this, we adopt the standard notation from the context free parsing literature that represents the beginning of a phrase by placing a dot in front of it, and the termination of a phrase by moving the dot to its right.¹⁰ Thus when the parser sees a verb in a phrase like (23a), it builds the representation (23b) and on encountering and attaching the following NP, and detecting no following words, it moves the dot over, deriving (23c).

- (23) a. ...believe John
 b. D(.VP, V)
 c. D(VP. V NP.)

We thus allow the parser to perform the following three actions:

- (24)a. Make assertions about material it has seen.
 b. Add to the assertion set after seeing disambiguating material either by adding a statement to the

assertion set or adding features defined by the X'conventions to incomplete X'predicates that appear in previous assertions after seeing disambiguating material.

c. Establish and add to linear precedence statements.

Obviously, we must place some bound on how long the parser can wait to commit itself to full analyses of lexical material and we must also bound how much previously encountered material a parser can use in order to make its decisions. Given the psycholinguistic evidence cited above, if we make no restrictions, the parser would not be plausible because we would commit it to holding unbounded stretches of previously encountered material since it would always allow domination statements to be added to its previous structure.

In order to put well tailored bounding conditions on the system, let us briefly recapitulate the basic features of the Marcus system that we discussed in chapter one:

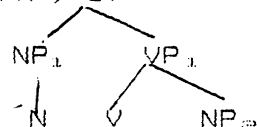
The Marcus system produces stacked representations on its active node stack. These elements will now consist of domination statements. Since parsing decisions can be based on previous domination information we will have to insure that each domination statement contains a bounded amount of information and that a parsing move (addition of domination statements, etc) can only make reference to a bounded number

of these domination statements. Let us solve the first problem by claiming that each domination statement can only contain information about what the parser currently claims are its immediate daughters. That is, a domination statement can be of the form (25b) but not (25c) for a sentence like 25.

(25)a ...believe Nory likes lawnmowing...

b. $D(DVP, S)$

c. $D(VP, S)$



For the moment, the restriction that a domination statement can only contain one level of embedding (as opposed to 3 or 22 levels) is a stipulation which we will justify in the next chapter.

Given this restriction, information about daughters of daughters of a category must be contained in separate domination statements. Proceeding with this example, we would add 25d and e to the set.

(25)d. $D(S, NP_1 VP_1)$

$D(VP_1, V NP_2)$

$D(NP_1, N)$

Besides associating a set of domination structures with a unique syntactic tree in those cases where speakers give the corresponding sentences a unique interpretation, the

parsing routine forces parsing decisions to be made based on c-commanding information only for the majority of cases. We will see in chapter 7 that in addition to making the model more psycholinguistically plausible, this architecture also correctly rules out a variety of unacceptable sentences.

In order to completely restrict the parser from using unbounded amounts of previously encoded material, we have limited the number of domination statements in a substack that can be accessed during the parsing process.

2.6 An Apparent Problem for Non-Backtracking Models:

In this section, we examine Frazier and Rayner's (1982) experiments and show that, while they are consistent with the minimal commitment theory, they are not consistent with a deterministic parser with a lookahead.

In fact, Frazier and Rayner claim that their results are inconsistent with both a deterministic processor with a lookahead and a parallel model. Frazier and Rayner measured subjects' eye movements as they read a variety of locally ambiguous sentences. The sentences were those like (1) above and (26)

(26) a. The lawyers think that his second wife will claim
the inheritance.

- b. His second wife will claim the inheritance belongs to her.
- c. Since Jay always jogs a mile seems like a very short distance to him.
- d. Since Jay always jogs a mile, this seems like a very short distance to him.¹¹

These sentences were presented without the disambiguating comma. The (b) sentences were uniformly more difficult to process. In the (b) cases subjects gazed for longer periods of time at the disambiguating phrase (the postnominal verb that signals that e.g. 'a mile' is a sentential subject). They also exhibited regresssive eye movements to the point of local ambiguity (the phrase 'a mile' in (26) that can either be attached to the adverbial or main clause. This pattern was only observed in the (b) type cases. Frazier and Rayner argued that these results were inconsistent with a parallel model because such a model, on encountering an ambiguous segment would pursue both analyses consistent with the material that it had previously seen. Therefore, the observed assymetry would be difficult to explain. Things look no better for deterministic parsing models with a lookahead. Given that the deterministic device of the Marcus(1980) variety cannot misparse any material, we would expect it to pause at a locally ambiguous segment in sentences like 26a or 26b because in both cases it must look

ahead in order to unambiguously decide how to proceed. However, longer gaze duration was not observed at the ambiguous position in either 26a. or 26b. We would also expect regressive eye movements following lookahead in both cases.

By contrast, if we assume that a backtracking parser always pursues one analysis preferentially¹², then we expect the observed asymmetries. We do not expect longer gaze durations at the ambiguous point of the structure because the parser does not resolve the ambiguity at this point in either case. Rather it executes the preferred analysis of the structure automatically. If the disambiguating material reveals that the nonpreferred analysis was, in fact the correct structure, then the parser must backtrack and recompute the analysis from the point where it made its initial error. This would comport with the regressive eye movement pattern observed in non-preferred readings like (26b).

Notice that it is the lookahead mechanism of the deterministic parser that yields the incorrect predictions. In order to make the determinism hypothesis consistent with this data we must re-examine whether we need a lookahead mechanism in this type of device. These cases are fully consistent with a minimal commitment processor. At the ambiguous point, the parser will simply add as many

statements to its assertion set as it can make with full confidence. If the parser reaches a disambiguating point that is consistent with the minimal analysis that it has computed it does not have to add to its assertion set and so this material will be treated like any other phrase. In cases like (26b) and (26d) the parser must add to its assertion set and so we would expect both longer gaze duration at this point. Regressive eye movements are interpreted as in the backtracking model except that, we interpret the parser as simply adding to its assertion set at this point rather than changing the attachment at the point of ambiguity.

2.7 Some Advantages of the Minimal Commitment Theory over Backtracking and Parallel Devices

Considering only the results in 2.7 might lead us to conclude that the Minimal Commitment Version of the deterministic model was merely a baroque simulation of the more standard backtracking solution of Frazier & Rayner. So far, the most that we can say for this theory is that the emendations proposed are motivated independently as shown in section V and are not simply accretions designed post-hoc to handle Frazier and Rayner's experimental results. In this section, I would like to strengthen these conclusions by showing that the minimal commitment strategy allows us to account for two other unrelated experimental results. The

results, from Crain and Fodor (1985) is unexplained given a parallel model, as was the Frazier & Rayner data. The second result is from Gorrell (1987) and is not consistent with Frazier & Rayner's backtracking solution.

Via. Fodor and Crain

Fodor and Crain (1985) conducted a self paced reading task on sentences like those in (27) and (28).

(27) a. Who could the little child have started to sing those stupid French songs for last Christmas?

b. Who could the little child have forced to sing those stupid French songs for Cheryl last Christmas.

c. Who could the little child have begged to sing those stupid French songs last Christmas.

d. Who could the little child have begged to sing those stupid French songs for last Christmas.

(28) a. The little child started to sing those stupid French songs for Cheryl last Christmas.

b. The little child forced us to sing those stupid French songs for Cheryl last Christmas.

c. The little child begged to sing those stupid French songs for Cheryl last Christmas.

- d. The little child begged us to sing those stupid French songs for Cheryl last Christmas.

They were interested in whether cases like 27 (c) and (d) would be more difficult to process than the other sentences in 27 or than their declarative counterparts, 28 (c) and (d). 27 (c) and (d) could conceivably be more difficult than the other sentences in this group because verbs like "to beg" are locally ambiguous in terms of the complement type that they select. As 28 (c) and (d) show 'beg' can subcategorise for either an NP S with object control into the S complement or for a simple S complement with subject control. By contrast, verbs like 'started' and 'forced' subcategorise uniquely for a sentential complement with a PRO subject and an NP, S complement with object controlled PRO, respectively.

The question is interesting because it can help us to choose between parallel and non parallel models. A parallel model predicts that both analyses consistent with (27 c and d) should be pursued in parallel and so neither of the underlined continuations of this sentence should be problematic. The structures are given in (29).

- (29) a. Who_i could the little girl have begged e_i [_{CP} PRO_i to sing those stupid French songs for Cheryl last Christmas.

- b. Who_i could the little child_j have begged [_{CP}Pro_j to sing those stupid Fr. songs for t_i last Christmas.

By contrast, a minimal commitment or backtracking theory predicts that 29a should be a more difficult sentence than any of the others in 27 and also more difficult than its matched declarative control. Within the minimal commitment framework, this is because the parser will pick the minimal hypothesis consistent with the input. Since it cannot be absolutely sure that the "WH" element should be taken as a complement to the verb 'beg' and since the postulation of a trace in a case-marked position that this analysis requires, would cause the parser to go into an error condition should the analysis be incorrect, the parser will delay postulation of this category. This means though, that if the parser reaches the end of the sentence without finding an alternative theta-position to interpret the WH element in, it should make the analysis increasingly complex and finally force the parser to go into an error condition. A backtracking parser that includes a minimal commitment heuristic will make the same predictions.

Crain & Fodor's results confirm the minimal commitment interpretation. Crain & Fodor measured the reading times for each word in their stimuli¹³ which they divided into 11

positions. We will be interested in their analyses for positions 7-11. Their analysis is:

<u>Position</u>	<u>Description</u>	<u>Examples</u>
1	Initial words	who could the little child have the little child could have
2	Main verb	start, force, beg
3	Potential gap position, sometimes filler (answer)	ti, us
4	embedded infinitive	to
5	complement verb	sing
6	NP	those stupid French songs
7	P	for
8	gap in nonminimal reading lexical NP in minimal reading	e _i Cheryl
9-10	Adverbial	
11	sentence final punctuation.	

Increasing complexity was measured by subtracting the reading time at position 7 (the position of the gap in the

minimal reading) from the sum of the reading times at position 9, and 10. For the non-minimal reading, the reading time at position 8 was subtracted from the sum of the reading times at position 8 and 9. Increasing complexity scores in both cases were also taken by subtracting the reading time at the final position from that of the previous position.¹⁴

High increase scores measure extra processing load and can be interpreted as showing that the parser was garden pathed at the position where these scores appeared.¹⁵

Fodor and Crain first looked to see whether the unambiguously minimal or non-minimal verbs ('start' and 'force') differed from their declarative controls in position 8. They found no significant difference in either case. This makes sense if we claim that the parser can use subcategorisation information (as we have been assuming) in forming its syntactic analysis of these structures. Both of these structures are disambiguated at position 5 by the subcategorisation information of the relevant verbs and so the parser will choose the right structures in both cases and not be garden pathed at later positions in the sentence. C&F found however that there was a significant difference between the non-minimal question reading of ambiguously non-minimal verbs like 'beg' and their declarative controls at position 8. This suggests that the parser first computes a

minimal reading (29b) ,which is consistent with the material up to position 8 for ambiguous verbs . At position 8 , it uncovers input that is inconsistent with the initial analysis and so it goes into an error condition because it must return to a domain that is more than two maximal projections away from the disambiguating material to add the object trace to the matrix sentence and bind the PRO to this trace as is shown in (29a).

At position 6, C&F found that the increase score for the unambiguously non-minimal verbs was higher than for that of the ambiguously non-minimal class. This reinforces the conclusion of the previous paragraph because it suggests that the parser computes the minimal reading for the ambiguous case (which involves creating only one empty category (the PRO) at this position and should thus take longer to process than the unambiguously non-minimal reading which involves creating both a wh-trace and embedded PRO subject in this position.¹⁴

As mentioned at the beginning of this section, these data are inconsistent with a parallel architecture because this model predicts that the parser would construct both readings consistent with the information up to position 8 for the ambiguous verbs. Therefore the parser should have already built an analysis compatible with the non-minimal reading revealed at position 8. Thus we do not expect a higher

increase score at this position for these cases, contrary to C&F's results. We also might expect a higher increase score at position 6 for the ambiguous cases when compared to unambiguously simple cases because the parser would have to create a reading with two empty positions at this position for the ambiguous verbs. C&F failed to find this result, they also did not find that these verbs patterned with the unambiguously non-minimal structures in this position. This result is also contrary to what is predicted by a parallel model because in both cases the parser is creating two empty categories at this position for at least one of the readings consistent with material at this position.

We should also note that these data are also consistent with a backtracking parser that preferred a subject control reading over object control for the ambiguous cases .¹⁷

Gorrell's Lexical Decision Task:

We will use the next set of data to choose between minimal commitment and backtracking models. The data come from Gorrell (1987). Gorrell used a lexical decision task that had been shown by Wright and Garrett (1984) to be sensitive to syntactic context. Wright and Garrett had shown that subjects are faster at lexical decision when the lexical decision target is the head of a maximal projection that is consistent with previous syntactic context than in the

control case where an inconsistent target is presented. The examples in (30) contrast targets that are syntactically appropriate (30a and b) versus inappropriate targets. The target item is capitalised in each case.

- (30) a. If your bicycle is stolen, you must FORMULATE
 *b. For now, the happy family lives with FORMULATE
 *c. If your bicycle is stolen, you must BATTERIES
 d. For now, the happy family lives with BATTERIES.

Wright and Garrett found that subjects responded to the target nouns and verbs significantly faster when they were placed in the appropriate syntactic context, than in the unacceptable controls.

Gorrell(1987) adapted this result so that it could be used to show what analyses subjects computed for verbs with ambiguous complement structures. He presented subjects with triples like those in (31)

- (31) a. Its obvious that Holmes saved the son of the
 banker/ right away.
 b. Its obvious that Holmes suspected the son of the
 banker/
 c. Its obvious that Holmes realized the son of the
 banker/ was guilty.¹⁹

The sentences in the triples contrast in that the first verb unambiguously takes a simple noun complement, the second verb can take either a simple noun complement or a sentential complement with a lexical subject, while the third verb can only take a sentential complement. At the point indicated by the "/" in (31), Gorrell presented subjects with either a set of verbs or pronouns as lexical decision targets. Neither target was appropriate for cases like (31a) as shown by 32a. The verb target was appropriate for the non-minimal reading in 31b (as shown by 32b vs. 32c.) The verb target was also appropriate as a continuation for 31c (as shown in 32d vs. e.). Items enclosed in brackets are the lexical decision targets.

- (32) a. *Its obvious that Holmes saved the son of the banker
[WAS, HE] right away.
- b. It obvious that Holmes suspected the son of the
banker[WAS]
- *c. Its obvious that Holmes suspected the son of the
banker [HE]
- d. Its obvious that Holmes realized the son of the
banker [WAS] guilty.
- *e. Its obvious that Holmes realized the son of the
banker [HE]...

If we assume, as we did to explain Frazier and Rayner's data, that subjects first compute the minimal reading of ambiguous structures, then since neither of the lexical decision targets is an appropriate continuation of either 31a. or 31b, neither target should benefit from facilitation by syntactic context. However, if subjects compute a structure consistent with the non-minimal reading of 31b, then we would expect facilitation for lexical decision on the verbal target since verbs (as opposed to the pronoun targets) serve as appropriate continuations of the non minimal reading. Thus this experiment can serve as a clear test of the backtracking hypothesis. In order to explain Frazier and Rayner's data, backtracking theorists must predict initial preference for the minimal reading in ambiguous cases, thus predicting that the ambiguous cases will pattern with the unambiguously minimal cases in Gorrell's lexical decision task because subjects must first backtrack, and revise their initial hypothesis in order to make the sentence consistent with the verbal target. At the time of verb target presentation, the ambiguous sentences should be inconsistent with this reading. Gorrell found however that the ambiguous cases pattern like the unambiguously non-minimal verbs in that lexical decision on the verbs was significantly faster than for the pronouns in 31b and c.¹⁹

Gorrell claims that his data argue that the natural language processor is a parallel device because they show that

subjects must compute both readings of the ambiguous verbs, with computation of the non-preferred, non-minimal reading accounting for the subjects' lexical decision judgements. This interpretation presents us with an unparsimonious picture because, as we argued above, the parallel model cannot account for the data from either Frazier and Rayner or Crain and Fodor's experiments. Obviously, we would prefer a model that could make sense out of the full range of experimental tasks.²⁹

We showed above that the minimal commitment theory predicted the data observed in the Frazier and Rayner experiment and was consistent with the Crain and Fodor data. This theory also predicts Gorrell's findings. This is because, the theory will build a parse that is consistent with verbal targets in both the ambiguous and unambiguously non-minimal cases. This is because, in the ambiguous 31b, the parser will have committed itself only to saying that the verb 'suspects' dominates the noun phrase 'the son of the banker'. Thus the parser builds the domination descriptions shown in 33.

(33) It's obvious that Holmes [_{VP}suspected]_S [_{NP}the son of the banker [_{VP} was]....

D(VP, NP)

It has not committed itself to saying that the verb directly dominates this phrase. The verb target is consistent with the non-minimal reading for the ambiguous case which the parser can compute by first adding the domination statement in (34a) and then in (34b) to the description in 33.

(34)a. $D(S, VP)$

b. $D(S, NP)$

Holmes <suspected <the <son <of <the <banker <was

As mentioned above, there is no consistent continuation for either lexical target for the simple verbs or for the pronoun target in either reading of the ambiguous case. Therefore we correctly predict facilitation effects for the verbal continuations in the ambiguous and complex (non-minimal) cases.²¹

2.8 Conclusion:

In this section, we have discussed three types of computational architectures: parallel, backtracking, and deterministic parsers. Within the deterministic class, we discussed parsers with and without lookahead and empirically justified a minimal commitment strategy that allowed the parser to handle natural language without recourse to

lookahead. We also presented psycholinguistic arguments to show that if the natural language parser is deterministic, it must have no lookahead capacity.

12. This could either be because an analysis is precoded into the parser's rule format or because one attachment strategy is independently motivated and follows from more basic principles of parser design as Frazier and Rayner (1982) and Frazier (1985) argue)

13. "In order to minimize the effects of different overall reading speeds for different subjects, all reading times were converted to z-scores...the standard deviation from that subjects overall means word reading time."

C & F pg. 118.

14. In all cases "reading time" should be read as "z-score" of the reading time.

15. These data are interesting in another respect in that they show that we should not generalise our empty operator strategy to all empty categories. We might imagine that we could insert traces as well as empty operators at will, assuming that they would simply remain uninterpreted if they were not associated with an operator. Thus in a case like (a) above, we would simply insert a trace in the matrix postverbal position which would remain uninterpreted if we happened to be dealing with a structure like (b).

- a. Who_i could the little girl have begged t [_{CP}PRO_i to sing those stupid French songs for Cheryl last Christmas.
- b. Who_i could the little child_j have begged t [_{CP}Pro_j to sing those stupid Fr. songs for t_i 1st Christmas.

The algorithm proposed above does not allow this because, since the unindexed trace in (b) is in a theta position and a case position it will be visible in both PF and LF and so cannot remain uninterpreted. Rather, it will cause the parser to go into an error condition because it will interpret the unindexed trace as a free variable which natural languages do not allow.

This seems to be the right conclusion, because, if we adopted the alternative algorithm, then we would not be able to explain why the parser garden paths in a case like (a). It would have already created a trace which could then be bound to the operator.

16. We should make two additional points to give a clear picture of the Crain & Fodor results. First, at position 8, C&F failed to find a significant difference between ambiguous and unambiguously non-minimal sentences (disregarding their

declarative controls). However they note that this does not undermine the garden path account of the ambiguous case discussed above because this lack of an effect could also result from the subjects' uncertainty about whether position 8 was the only remaining potential gap position in these structures. That is, a minimal reading for these sentences predicts a WH trace position after position 6 in all cases but subjects, upon realising that position 8 was lexically filled, may have concluded that this trace position would appear at a later point in the sentence. This predicts that the increase score for the final position should have been higher for cases in which the nonminimal reading was the correct one for an ambiguous structure than when the parser was computing an unambiguously non-minimal reading which is what C&F observed. (See C&F pg 122).

We should also note that C&F did not find a significant difference between the increase scores at position 6 for the unambiguously minimal and non minimal verbs. Thus without the data from position 8, we could not be as confident that difference between ambiguous and unambiguously nonminimal verbs could be related to the complexity of the analysis at position 6.

17. This strategy might be motivated by the fact that subject control seems to be the unmarked option for natural language in the sense that there are many languages that allow only subject control structures while there are no corresponding languages that allow only object control.

18. See Gorrell (1987) pg.28.

19. The data for the relevant experiment is given below:

"Mean response times and percent correct for context/target pairs ..."

TARGET

<u>Context</u>	<u>Verb</u>	<u>Pronoun</u>
Simple	771 (96%)	753 (98%)
Ambiguous	690 (95%)	762 (86%)
Complex	696 (97%)	761 (94%)

Verbal targets used in the experiment were 'is, has, must' versus pronominal targets 'me, him, them'.

'Simple' corresponds to 'unambiguously minimal' verbs (verbs that unambiguously take simple nominal complements.

'Ambiguous' corresponds to verbs that can take nominal or sentential complements and 'complex' corresponds to verbs that only take sentential complements.

Gorrell (1978) pg. 40.

20. In subsequent experiments, Gorrell also finds that ambiguous verbs pattern with the unambiguously non-minimal group. However, since these experiments asked subjects to judge the grammaticality of sentences that had just been presented to them, Gorrell correctly argues that this pattern could result from the fact that since the minimal reading takes less time to compute (it contains fewer constituents than the non-minimal reading) it would be the first one submitted to the post syntactic decision making component of the processing device. Thus even if the syntactic component computed all possible structure in parallel, we would expect these data because the judgement task calls upon a post syntactic, decision making process. We chose to discuss the Frazier and Rayner and Crain and Fodor experiments because the same argument cannot be made in these cases. That is, these authors have argued extensively that eyetracking and self paced reading reflect the workings of the on-line syntactic component, and do not involve the post syntactic decision maker. Preliminary support from this claim comes from simply noticing the fact that reaction times for these tasks are significantly faster than for grammaticality tasks. See also Frazier and Rayner (1982) and Crain and Fodor (1985) for discussion of the on-line nature of their tasks.

21. In the main text, we have actually analyzed half of the results that Gorrell presents in favor of the parallel processing hypothesis. The other half of his results have to do with the ambiguities between main verb and participial presented by cases like (a):

- (a)' The old man was shaved/whenever he was was ill.
- '' The old man shaved in bed/
- ''' The old man they shaved in bed.

The interesting case is a'' which Gorrell notes is ambiguous between a "main verb" reading as in (b') or a participial reading as in b''.

- (b)' The old man shaved in bed
- (b)'' [NP the old man_i [_S O_i [_S e_i shaved e_i in bed]]]

As in the experiments discussed above, subjects were presented with either verbal or pronominal targets at the point indicated by the "/". Also, as in the first experiment, subjects responded significantly faster to the verbal targets when they followed unambiguous relative clauses like a''' or cases that are possibly relative clauses in their non-preferred readings like a''. Again, Gorrell argues that this suggests that subjects must be computing

both the preferred and non-preferred readings for the ambiguous cases.

While this analysis will explain his data, it leaves us with no explanation for why this reading is so inaccessible to conscious interpretation. If both readings are computed, then both will eventually be submitted to the post-syntactic analyzer, with the non preferred reading eventually winning out over the preferred reading as the parser encounters a verb that disambiguates the structure in favor of the non-preferred reading.

However, it seems that this effect may be an artifact of the particular class of verbs chosen for the experiment. Gorrell presents 18 verbs that can be ambiguous between the main verb/participial reading. Of these 18 verbs, 8 are actually put in frames like a'' and followed by a verbal target. However, three of these eight verbs (given in the sentence fragments used by Gorrell in his experiments in (c)) are ergatives as shown by the fact that they can be used in the stative passives or causative constructions in (d). Manzini (1982) shows that both of these constructions can only be formed in English from verbs with underlying objects.

(c)' The prisoners_i marched e_i along the road.

'' The oil_i poured e_i onto the road.

''' The soup_i boiled e_i in the pot.

(d) ' The general marched the prisoners down the road.

'' The oil is all poured from the can and is ruining the driveway.

''' The soup is already boiled and shouldn't be cooked anymore.

Assuming this analysis, the parser might initially treat the fragments in (b) as simple ergative structures. The presence of the target verb 'has' would not be inconsistent with this analysis if the initial fragments were reanalysed as sentential subjects.

Notice that, unlike the case b'' this analysis does not force any undoing of the initial thematic role assignment internal to the simple ergative structure. The only reason that the sentential subject analysis cannot be maintained is that it violates the *NP Tns VP filter. This information is arguably not used by the parser as it constructs its initial analysis. Eventually, the parser will have to restructure this analysis into a relative clause with an NP head, empty operator, and empty trace in the subject position of the relative clause, but even these additions will not violate the determinism assumption because they do not involve revocation of any decisions about the case of theta structure of these initial clauses. In addition, ergatives are like the passive participals in the reduced relative reading in

Chapter 3: Subjacency and the Deterministic Model

In this chapter I use some linguistic and computational arguments to buttress the psycholinguistic evidence for a bounded context deterministic model of language comprehension and discuss the constraints that such a parser imposes on its input grammar.

In the previous chapter, I examined the constraints placed on the design of the human parsing algorithm by the problem of encoding grammatically relevant information in a psycholinguistically verified and computationally tractable way. We can approach the problem from a different perspective as well; namely we can ask what conditions the design of efficient parsing algorithms imposes on principles of grammar. In Berwick and Weinberg 1984 we tried to show that the Subjacency Condition, an axiom of the Government Binding Theory had a functional motivation in terms of efficient parsers. That is, we tried to show that the subjacency condition was invoked by a deterministic parser in order to provide local disambiguation for cases that would otherwise be locally ambiguous and that local disambiguation was crucial for guaranteeing efficient parsability in these cases. We tried to provide both conceptual and empirical motivation. Namely, we are not claiming that it is logically necessary that the parser be able to process all sentences

generated by the grammar efficiently, or even at all. Rather, we notice a class of locally ambiguous cases that the parser seems to be able to disambiguate without exhibiting a preference for one reading over another. We therefore conclude that the parser must be able to disambiguate these structures on line. We further notice that disambiguation entails access to previously encountered context and so we conclude that the parser must be able to access this information efficiently. The idea that this condition has a functional motivation is reinforced by looking at the cases where it applies and noticing that this condition applies whenever the parser would need to access left context in order to build a syntactic structure correctly. In this chapter, I will briefly review the relevant results from Berwick and Weinberg 1984 and extend the treatment provided there in a way that explains the curious set of construction types that are governed by this condition. The set is curious from the grammatical perspective because the cases that are governed by Subjacency do not constitute a natural class. This condition cannot be seen as a natural condition on movement because while it governs both WH movement, and NP movement, it also governs the parasitic gap construction, and a subset of gapping cases which involve no movement at all. It does not govern LF movement. The condition is also irrelevant to Binding, or Control. We will see that subjacency stands as the only condition on S-structure

representations. Even when viewed as a condition on representations, the condition is unnatural because it applies only to cases where empty categories are involved and only to a subset of these constructions. If the condition had no motivation external to the grammar, we would expect it to apply to a class of rules that the grammar construed as natural. We might expect the condition to govern the output of either all movement, all movement and deletion transformations or to all S- structure gaps. We will see however that the full set of constructions governed by Subjacency form a natural class in that they present an efficient parser of a certain type with a unique computational problem. Imposing a ban on unbounded linking between an antecedent and the category it binds allows the parser to find a unique solution to the problem presented by all of these constructions. This chapter concentrates on S- structure phenomena and introduces the way in which subjacency governs the parasitic gap construction,¹ and gapping constructions. We explain both why these structures should be governed by subjacency and licensed at S-structure. We also show that subjacency does not govern LF movement. Finally, we discuss what it means for the parser to impose a condition on a grammatical formalism and how this condition is enforced by Universal Grammar.

This account is interesting in two other respects. First, we will see that the basic parser design that we motivated in the last chapter using completely unrelated data plays a crucial role in resolving the aforementioned puzzles. In addition, we will see that the features of our account are all independently motivated and inconsistent with truly nondeterministic parsers. Therefore this kind of evidence allows us to eliminate a rather large class of mechanisms from the set of "possible parsers" underlying human language comprehension, a major goal of this thesis.

3.2 Bounded Context Parsing

In this section we introduce the idea of bounded context LR(k) parsing and discuss the basic structure and efficiency of this class of processing devices. LR(k) parsers have the following properties.¹²

- (1) LR(k) parsers are deterministic in the sense discussed in the previous chapter.

That is, an LR(k) parser must always be able to unambiguously make its next move based on what it can store from material that it has previously encountered and a finite amount of lookahead. If we think of this condition in terms of parsing¹³ a simple phrase structure grammar, this means that

the parser must be able to look at the structure it has already built, and a bounded number of lookahead tokens, and decide how to expand the phrase structure rule that it is currently building."

- (2) Previously encountered material must be representable as a finite control table.

This places restrictions on the type of information that the parser can use to make its decisions about what to do next. As is well known there can be an unbounded distance between the WH-operator and the variable position on the surface in questions.

- (3a) Who do you believe that Mary said that Fred decided to visit?

One might represent the material between the operator and the trace position as variable in the form of a rule like (3b)

- (3b) WH X -----> WH X trace

where the 'X' stands for an unbounded stretch of terminal and non terminal material. This encoding is unavailable in the LR(k) framework because it cannot be represented in the device's finite control table. A finite control table must

explicitly encode a set of actual terminal and non terminals in its stored set of grammar rules.³ Thus if the parser has to look at material that is unboundedly far away from a state in which it has to make some parsing decision, this material has to be recoded so that this information can be recaptured at a place that is finitely distant from the state at which it will be needed to make a parsing decision. The alternatives open to an LR(k) device for questions are discussed below.

- (4) The parser can use only a finite number (k) of lookahead tokens in order to make a parsing decision.

Given the psycholinguistic evidence cited in the previous chapter, we will place the most restrictive requirement possible within this framework, limiting the parser to using only material it has previously seen and the current input token that it has uncovered. That is, we claim that a psycholinguistically motivated parser is an LR(0) parser. As Knuth (1965) and many other authors have noted, there is a rather substantial payoff for placing these restrictions on the parser's architecture. Knuth(1965) proved that if a language can be parsed by an LR(k) parser, we can guarantee that it will be able to parse sentences in time that is linearly proportional to their length. That is, if we can show that sentences of natural language can be parsed in the

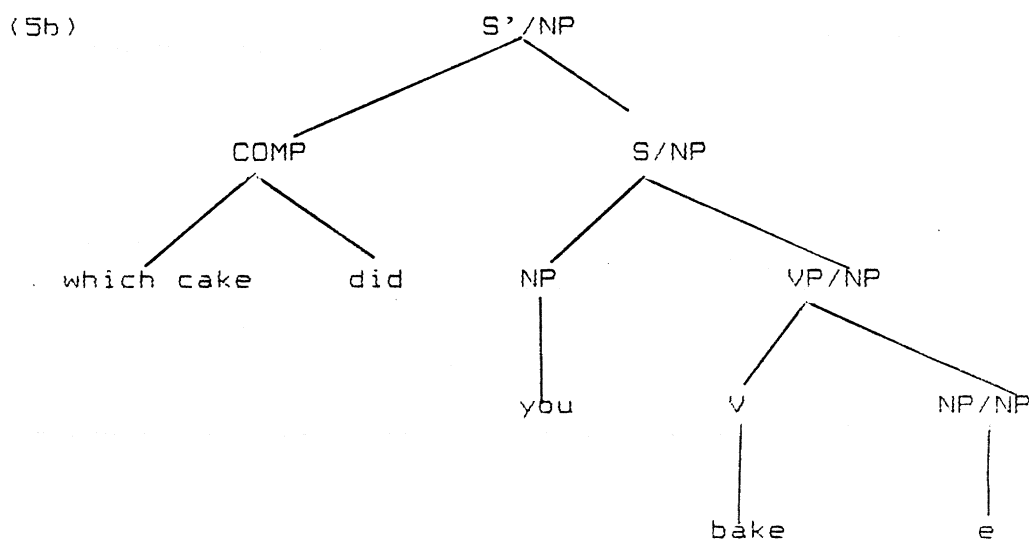
LR(k) framework, we can model the ability of human beings to understand the majority of sentences in their language in basically the time that it takes to read or utter the words of those sentences. That is, we can guarantee that the process of associating the surface string with a deep structure does not add unduly to the language comprehension time.

The force of the argument presented in this chapter depends on our being able to show that Government Binding theory is naturally implemented by only a subset of LR(k) parsers; the bounded context parsers. This restricts the class of LR(k) parsers to those where the left context must be "literally" finitely encoded in the sense that we must insure that there will always be a finite number of non terminal symbols intervening between the symbol that the parser needs to retain in the left context to make a syntactic decision and the point in the tree where the decision is made. The ban on essential variables does not insure literal encoding of left context because as we showed in the previous chapter, it is possible to obtain the effect of essential variables without actually using them. One does this by using the method of generative encoding.

The first method discussed was the "slashed category" notation from GPSG. Recall that this notation allows the

grammar/parser to pass down the presence of a WH element through the phrase structure so that its presence is recoverable at any point by reference to the immediately dominating phrase. A sentence like (5a) has the representation (5b).

(5a) Which cake did you bake?



The price of this choice lies in the expansion of the number of phrase structure categories. Such a theory now needs those PS categories sanctioned by X' theory, and also must add a set of categories whose sole purpose is to locally encode the presence of previously viewed fillers. As shown in Berwick and Weinberg (1984), and Barton, Berwick and Ristad (1987), this method exponentially expands the size of the grammar. Since, under a rule based system, the parser

must search through the grammar space in order to find the appropriate rule for the sentence under analysis, this procedure will greatly impede our goal of efficient parsability. Moreover, this representation is not consistent with the view that phrase structure rules are merely the projections of lexical categories. In such a theory, the only categories that are available for the construction of phrase structure are those that are independently generated by X' theory, i.e. Heads and their projections. Furthermore, the X' system does not in general distinguish between projections of empty and lexical heads in the way necessary for the interpretation of these "slashed" categories.

Generative encoding can also be achieved by augmenting the parsing mechanism. An example is the ATN system discussed in chapter 2. The HOLD CELL in that system provides the parser with a way of locally checking whether a WH element is available to fill a gap position at any point in the parse. Since the HOLD CELL removes the WH element from the rest of the left context representation, the parser is never forced to wind back through unbounded stretches of left context. The costs of this approach were reviewed in chapter 2 and so we will not dwell on them here.

Let us adopt a transparent representation of a GB grammar as a parser. Since both methods of generative encoding involve

unmotivated additions either to the rule writing vocabulary or to the parsing mechanism, we must insure local encoding of left context by using the class of bounded context LR(k) parsers that allow only a literally finite encoding of left context.

3.3 The Problem of Ambiguity and Left Context for Bounded Context Parsers

As noted in the previous chapter, the main problems facing a deterministic parser are those where the grammar presents the device with multiple alternatives at a given point in the parse. Consider a sentence like (6) in this light.

(6) What₁ do you think Fred ate e₁?

To analyze this sentence correctly the parser must insert an empty category after the verb. Notice that this verb has an ambiguous subcategorisation structure and takes an intransitive reading as well.

(7) Do you think Fred ate?

There are three possible ways for a parser to deal with this problem. It could simply output one of the

representations consistent with the locally available subcategorisation information. This would predict that it should be difficult for the parser to process one of the readings given above correctly, which does not seem to be the case. Secondly, we could give up the assumption of determinism, resorting to a parallel model in which both structures would be generated, but where only one would ultimately receive an interpretation. Such an account would have to distinguish the cases discussed in chapter 2, where the parser seems to compute a single reading for multiply ambiguous structures, from cases like (6) and (7).⁴ The final possibility is for a bounded context parser to introduce an algorithm that guarantees that the information it needs to locally disambiguate a structure will be present within its bounded context.

A deterministic answer to the question of whether to insert an empty category after the verb, cannot be gotten simply from the analysis of the local subcategorisation environment. Rather, the only firm evidence for the need for the expansion of an empty category in (6) is the presence of the WH operator at the beginning of the sentence. The problem however is that, on the surface at least, the disambiguating operator can be indefinitely far away from the site of the potential empty category as the example in (8) shows.

- (8) What_i do you believe Mary suggested Fred thought that Bill said that Frank believed that John expected Sue to eat e_i

This suggests that if the parser had to consult the WH element at the beginning of the sentence directly, it would have to store a potentially unbounded left context. This is because we cannot guarantee that a finite number of nonterminal symbols will intervene between the trace and the antecedent if we use an actual encoding of the intervening non terminal symbols between the antecedent and its trace. This entails that some bound must be placed on the distance between the empty category and the element that binds it. If we assume successive cyclic movement and Subjacency, we can guarantee that the parser will only need a finite amount of left context in order to make its decisions in these cases. Assuming the parsing analogue of successive cyclic movement the decision to insert a trace into a structure like (9) below can be made on the basis of the presence or absence of a trace in the COMP of the adjacent S. This guarantees that we will only need to use a finite left context (the limit being placed by the upper bound on how many nonterminals one clause can contain) in order to make the decision about whether to expand the trace.

- (9) What_i do you believe [_S e_i [_S Mary suggested... [_S e_i that Sue eat...]

If there is a trace in the subjacent COMP, we can expand the phrase structure with a trace, if not, we don't.⁷ Notice however that any bound on the distance between the antecedent and trace will be sufficient for our purposes. One must then wonder why the condition is one of sub as opposed to tri jacency. Berwick and Weinberg 1984, suggested that we were limited to sub jacency because the parser reflected the grammar in that the predicates used in its rule writing formalism could not include counting predicates. Therefore the parser had to use a bounding predicate that could be stated in non numerical terms. The only case where numerical notions can be stated using non numerical predicates is the predicate for 'one away' which we can reformulate as 'adjacent' or 'next to'. Therefore the fact that bounding is expressed as Subjacency follows straightforwardly.

We might also wonder why this condition applies to all cases of WH Movement and not only to the cases where the detection of a trace position involves retrieval of elements in the left context. Alongside cases like (6) we have cases like (10) involving an obligatorily transitive verb.

- (10) Who did you hit?

Again we appeal to a condition of "natural implementation". The rules of the grammar are not permitted to be written using either Boolean predicates or existential quantification. But in order to tell the parser to look at left context only if the verb has an ambiguous subcategorisation frame we would need to use existential quantification. Thus, in order to write a deterministic trace expansion rule we have to allow the parser to look at left context in all cases. This gives us the Subjacency Condition for all cases of WH Movement.■

Interestingly, the link between parsing ambiguity and syntactic bounding inherent in this account seems to correctly predict the cases that are governed by subjacency. This account predicts that constructions that present the parser with syntactic ambiguities that cannot be disentangled from local context should be governed by subjacency, while those that do not create these problems should not be subject to this constraint. Notice that cases which involve the binding of phonetically present anaphors, pronouns, and names to antecedents are not constrained by subjacency, as shown by the examples in (11). This follows from our account because, since all of the potential bindees in (11) announce their presence by being phonetically pronounced, the parser does

and then carry out a post syntactic search of the previously built syntactic tree to discover the appropriate structure. Notice that the grammaticality of the structures in (11) suggests that the bounding condition imposed on the parser is imposed at the tree building, as opposed to the post syntactic binding, stage. The difference between empty categories and the overt class of bindees in (11) is that empty categories force the parser to look at left context in order to decide how to build the syntactic tree, while the cases in (11) only require the parser to rescan previously built structure in order to decide the binding possibilities for categories in this structure. Thus, even though it is possible to claim that in the parallel case, locality conditions apply to the post syntactic search phase, this approach would make the wrong empirical predictions because it would incorrectly assimilate cases like (7) with those of (11) and would predict that either both cases or neither of these cases was governed by subjacency.

If this account is correct it also suggests that the parser does not use the type of generative encoding mechanism embodied in the HOLD hypothesis. Recall that this approach involves supplementing the parser's architecture with a mechanism that allows it to store the presence of a WH in such a way that it can be reaccessed at arbitrary points in the parse. Supplementing the parser with this type of

mechanism predicts that we should be able to reaccess the presence of a WH element even inside a syntactic island. Notice that we could always put special stipulations on the HOLD mechanism to bar access to a stored WH category as the parser built a syntactic island, but this restriction would clearly be ad hoc and would be added only to model syntactic island effects.

Similar remarks apply to the "slashed category" mechanism discussed above. If any category can have a duplicate "slashed" notation, then there would be no reason to expect ungrammaticality to result from movement out of an island. We could always supplement the grammar with restrictions that bar passing a slash into a syntactic island, but these assumptions would be ad hoc.

3.4 Parasitic Gaps

This analysis predicts that the distribution of parasitic gaps should be governed by the Subjacency Condition, because as Fodor (1985) has pointed out, the same ambiguities are found in the case of parasitic gaps.¹⁰ Thus in (12) we can put a parasitic gap into the position following the adjunct's verb while in (13) this is not allowed.

(12) What_i did you cook e_i without PRO eating e_i.

(13) Can you cook without eating?

Chomsky (1986) provides the following examples which show that these constructions are in fact subject to the Subjacency Constraint as our theory would predict.

(14) Who_i did you read a book about e_i to e_i?

(15) Which man_i did you interview e_i without reading up on e_i?

(16) *Which man_i did you interview e_i without reading [_{NP}[the file]_j [_{CP}[_S you made e_j on e_i]]

In (14), both gaps are subjacent both from the complementiser and from each other. This is shown by the fact that both (17) and (18), with overt movement from both the parasitic and regular gap position are acceptable.

(17) Who_i did you read a book about e_i

(18) Who_i did you read the book (that Mary bought yesterday) to e_i.

Chomsky uses the contrast in (15) and (16) to argue that parasitic gaps are bound to empty operators and are licit only if they are subjacent to these operators or to the operator traces in COMP. These empty operators must appear at the head of the adjunct clause because they must be subjacent to the real gap chain. Put in terms of our parsing model, we can use the presence of the overt operator to signal the presence of the "real" gap.¹¹ The presence of the empty operator can, in turn, be used to signal the presence of the parasitic gap if it is in a subjacent position. For simplicity we will assume the theory of bounding nodes in Chomsky (1973) for this discussion but we will refine the substantive theory of bounding nodes below.

This analysis predicts that (16) is bad because, since the COMP of the relative clause is already filled by the operator that is linked to the NP's head, the parasitic gap operator cannot appear in a position that makes it subjacent to the real operator; see (19) below.¹² This contrasts with (15) where every trace is subjacent to the operator that licenses it, as shown in (20).

*(19) Which man_i [_S did you [_{VP} interview e_i] [_{PP} without [_{OP}_i
 [_S PRO reading [_{NP} the file]_j [_S that you made e_j on
 e_i]]]]]]

(20) Who_i [_Sdid you [_{OP} interview e_i] [_S without [_S OP_i [_{PRO} reading up on e_i]]]]?

We showed that the creation of empty operators causes no problems for a deterministic system in chapter 2. We can use their presence to license parasitic gaps in the appropriate structures. Thus we can make the parsing model predict the properties of this construction in a straightforward and independently motivated way. It is important to note at this point that we are not changing assumptions in an ad hoc way simply to model the facts.

The algorithm introduced and independently motivated in chapter 2 allowed us to create an empty operator if a potential antecedent was available in a c-commanding domain, without fear of causing a subsequent garden path. This is because, assuming that the empty operator had no subsequent variable position to bind, it would not be visible in LF. It also has no phonetic content so is not visible in PF. Therefore it is simply uninterpreted if it does not form part of an open sentence. With this algorithm in mind, let us consider cases (21) and (22):

(21) a. who did you meet without ...

(22) a. did you watch the movie without eating

In a case like (21) the parser recognizes that part of the sentence is an adjunct phrase. This signals the possibility of a parasitic gap appearing in the subsequent structure. Therefore the parser inserts an empty operator in the COMP position as shown below.

(21) b. Who_i did you meet e_i without [_S OP_j...

If the parser subsequently finds a gap position in a subjacent domain it can create a trace and bind the operator to it, thus associating the operator with a theta position as shown in (21c).

(21) c. Who_i did you meet e_i without [OP_j [_S PRO
greeting e_j.]

Compare this case with (22). In (22) the parser will also detect an adjunct and so create an empty operator. In this structure there is no c-commanding antecedent for an empty operator therefore no parasitic gap chain will be initiated within the adjunct and so 'eating' will be given an intransitive interpretation.

(22) b. Did you watch the movie without [_S [_SPro eating]].

In cases like (13) above, the adjunct again triggers the creation of an empty operator. Since there is no gap in the adjunct phrase, the operator is not associated with a theta role. Therefore, even though there is an overt operator to link with, the empty operator does not meet the criterion for visibility at LF and so is not interpreted. Since empty operators are not interpreted unless the condition on visibility is met, a deterministic parser can always create these categories because they can never cause it to simulate non-determinism either by backtracking or pseudo parallelism in order to correct past mistakes. Notice that this solution will only work for empty operators. Lexically specified elements will receive a phonetic interpretation but no semantic interpretation, a situation that will lead to unacceptability. An empty element with no semantic features however is neither semantically nor phonetically interpreted and so simply plays no role in the interpretation of the sentence.

Similar remarks apply to the empty gap position. The astute reader will have noticed an apparent problem created by this solution. Why, we might ask, if empty categories can become invisible at later stages of the interpretation, must we cue their creation to the presence of overt operators. The cases that motivated the account in the first place were those in

which the local subcategorisation of a verb was indeterminate. Before positioning an empty element after such a verb we claimed that we had to make sure that an actual operator was present in the previously analysed structure. However, given our present approach, one might be tempted to argue that if a verb that can be optionally transitive turns out to be used intransitively in a given structure, the gap will simply not be associated with an operator and so will become invisible at later levels of representation. This seems to remove the motivation for putting restrictions on left context which was crucial to the functional motivation for the condition in the first place. Recall however that only elements with no phonetic features can escape causing unacceptability if they are not semantically interpreted. Since WH elements have Case features they will be visible in the phonological component.

3.5 Licensing of Parasitic Gaps

This analysis has tacitly assumed that the real gap in a parasitic gap structure does not c-command the adjunct that contains both the empty operator and parasitic gap.¹³ The assumption has important consequences because, if we can justify it we can derive the otherwise stipulated fact that parasitic gaps must be licensed at S-structure. That is, we derive the fact that quantifiers and WH operators that move

to COMP or some other pre-S position at LF do not create acceptable parasitic gap structures as shown by the examples in (23).

(23) a. *_S[_SYou [_{OP} met who_i] [_{PP}without greeting e_i]]

b. *_S[_SEveryone [_{OP} met someone_i][_{PP} without greeting e_i]]

We know for reasons independent of bounding theory, that parasitic gap constructions are not licit if the real gap occurs in a c-commanding position like the Subject position in (24).¹⁴

(24) a. *Which man_i t_i got drunk before we saw e_i

b. *Everyman_i t_i got drunk before we saw e_i

As suggested above, the subjacency condition incorporates a c-command requirement. Chapter 7 gives some empirical motivation for this proposal. Neither the WH element nor the quantifier in (23) c-commands the adjuncts containing the parasitic gaps. Since the parser only uses c-commanding subjacent NPs or their traces to direct its creation of further trace NPs, there will be no evidence to create an empty operator in the COMPs of these adjuncts and thus no parasitic gap chain will be formed. Given that the input for

parsing decisions is the S-structure of the sentence, the subsequent movement of a category to a c-commanding position at a post S-structure level cannot help the processor decide how to expand the parse tree. Therefore our parsing theory can derive both the fact that subjacency is an S-structure property and the fact that parasitic gaps are governed by subjacency in that they must be licensed at S-structure -- the two central properties of the construction.

Recently, Heles Contreras has brought together some interesting evidence that might lead us to question our crucial assumption. Contreras (1985) claims that the real gaps in parasitic gap structures may c-command the adjuncts containing parasitic gaps. In the next section, we present some evidence to justify our original claim of non c-command between real and parasitic gap. We also critically evaluate Contreras' counterevidence.

3.6 On the Distribution of Epithets

In this section we will discuss the distribution of epithets. By "epithet", we mean a noun phrase with a pejorative connotation. We can informally describe the distribution of these phrases as being both like pronouns and like names. Like pronouns, they can be understood as coreferential to

another noun phrase in the sentence or discourse. Examples are given in (25).

(25) a. Mary knew Bill_i but [the bastard]_i wouldn't speak
to her at the party

b. Sue told Frank_i off before [the cretin]_i could make
excuses for himself_i

They also share the pronominal property of being able to act as a bound variable.¹⁵ By "bound variable", we mean that the epithet or pronoun can be linked to a c-commanding quantifier, and has its range specified by being so linked.

(26) a. Mary buttonholed [every senator]_i before [the
bastard]_i could get away

b. Mary buttonholed [every senator]_i before he_i could
get away

We can capture this resemblance by claiming that epithets have the status of pronouns. This would have the syntactic consequence that epithets should elicit "Leftness" violations in the sense of Higginbotham (1980).

This predicts that an epithet, like a pronoun, cannot be coreferential to a quantifier if it appears to the quantifier's (or its variable's) left:

- (27) a. *[[His mother]_i thought that [every contral]_i should get an F-15.

[the bastard]_i's mother

- b. *John_i knew [the bastard]_i thought that [every contral]_i should get on F-15.¹⁶

This cannot be the whole story though, because epithets are distinct from pronouns, and like names or definite descriptions in that they cannot be bound by a c-commanding NP anywhere in the sentence. Thus (28a-b) contrasts with (28c), which contains a bound pronoun, and with (28d) where the name or epithet is bound to a non-c-commanding NP.

- (28) a. *John [_{NP} persuaded him_i [_S that Bill_i was a fool]]
- b. *John [_{NP} persuaded him_i [_S that [the poor bastard]_i was a fool]]
- c. John [_{NP} persuaded Bill_i [_S that he_i was a fool]]

(28)(continued)

d. [_{NP}The man [_Sthat Bill_i trusted]] persuaded Mary
that [the poor bastard]_i was a fool
Bill_i

We can capture the entire distribution of this category by claiming that epithets fall under Principle C of the Binding Theory. Principle C claims that

An R-expression (name or variable (ASW)) may not be bound.

Thus we formally characterise epithets as both like pronominals and like names.

These facts about epithets, and particularly the fact that they cannot be bound to a c-commanding NP, can give us a sensitive probe into the c-command properties of parasitic gap structures. Recall that our initial question was whether the real gap in a structure like (29) c-commands the adjunct containing the empty operator or parasitic gap.

(29) Who_i did John criticize e_i in private while praising e_i in public.¹⁷

We can answer this question by placing a name in the position of the real gap and an epithet in the position of the parasitic gap. If the real gap position c-commands the adjunct and the parasitic gap, we predict the sentence to be ungrammatical. The grammaticality of (30a-c) shows that the real gap does not c-command the parasitic gap in (29) and that the epithet is functioning both as a pronominal and as a name, as our analysis predicts.

- (30) a. John criticized [the senator]_i in private while praising [the bastard]_i in public.
- b. John criticized [every senator] in private while praising [the bastard]_i in public.¹⁹
- c. [Which senator]_i did John criticize e_i in private while praising [the bastard]_i in public.

The data from epithets thus provide the crucial support we needed to buttress our derivation of the fact that licensing of parasitic gaps can only occur at S-structure by showing us that the real gap position does not c-command the adjunct containing the parasitic gap.¹⁹

3.7 Critical Remarks

In this section, I want to show that the data cited in Contreras 1985 should not be accounted for by claiming that the real gap c-commands the parasitic gap in adjunct structures.²⁰ Contreras (1985) notices that there cannot be coreference between the underlined pronoun and name in (31) when the sentence is pronounced with normal intonation.

(31) *John filed them without reading [Mary's articles]_i²¹

He argues that we could rule out this sentence if we claimed that there was a c-command relationship between these two phrases. The R-expression would be bound which would violate Principle C of the Binding theory.

This seems too weak though because we incorrectly predict that (32), where the pronoun is dominated by a PP and so does not c-command the name should be acceptable.

(32) *John talked about them_i to Fred without reading
[Mary's articles]_i.

We might be tempted to claim that the ungrammaticality here results from reanalysis creating a structure like (33)

(33) John [_{VP} [_{VP} talked about] them to Mary....

If this were the appropriate structure though, we would expect (34) to be grammatical because the reflexive would be c-commanded and bound. The ungrammaticality of (34) shows us that this cannot be the right structure.

- (34) *I talked about Thmug to himself
(This sentence is from Jackendoff)

Contreras' theory is also too strong in some of its implications. Contreras (citing an anonymous LI reviewer) remarks that a sentence like (31) sounds perfect with the altered stress pattern of (35).

- (35) John filed them_i without (even) READING [Mary's articles]_i.

He argues that this sentence is the result of right dislocation on a structure like (31) which results in a structure like (36) where the pronoun does not c-command the name.

- (36) John [_{NP} filed them_i [_{PP} without [_S PRO reading e_i]] [Mary's articles]_i.

It is unexplained and peculiar to have stress on the verb trigger right dislocation of a following NP. Moreover, it

seems that some kind of special stress pattern is associated with coreference even for the standard cases of backwards pronominalization. Many speakers do not get coreference in (37a) if the sentence is spoken with a normal intonation pattern. Backwards coreference is possible with emphatic stress on the preceding verb as shown in (37b).

- (37) a. His_i mother loves John_i.
 b. His_i mother LOVES John_i.

Given Contreras' analysis, (37b) should also be formed by right dislocation. The problem is though that if (37b) is formed by right dislocation, then (38) should also be formed by right dislocation. If this is true though, we incorrectly predict that (38) violates the ECP and should be unacceptable.

- (38) a. His mother THINKS John is nice.
 b. His mother [_{CP} [_{CP} THINKS [_S e_i is nice]] John]

Thus Contreras' account is too weak in that we can find cases where there is no c-command relation between pronoun and name, but where backwards pronominalization is nonetheless impossible. It is also too strong in that it links emphatic stress to right dislocation in a way that forces us to incorrectly rule out grammatical sentences. We have

obviously not given a complete account of the conditions under which backwards pronominalization is possible, but have suggested that the data Contreras was concerned with should be not be ruled out by c-command restrictions. By contrast, the distribution of epithets seems to be more straightforwardly ruled out by purely structural factors, and this probe tells us unequivocally that the adjunct is not c-commanded by the real gap position in parasitic gap structures. Given this, we derive the S-structure licensing of parasitic gaps.

3.8 Gapping Constructions

Fodor (1985) has noticed that "gapping" is not a unitary phenomenon. Rather she claims that gapping constructions are really the result of two distinct rules: "Main Verb Deletion" and "Tense Deletion". Interestingly, bounding constraints are observed only for Tense Deletion. We can show that our analysis can predict this distinction with no ad hoc complications.

3.8.1 Description of the Problem

Fodor (1985) points out that escape from bounding constraints correlates with the appearance of an auxiliary marker in the pregap position. (39) and (40) illustrate this contrast.

- (39) a. Mary fishes in the Ocean and Harry in the sea.
 b. *Mary fishes in the Ocean and I think Harry in the sea.
- (40) a. Mary has fished in the ocean and Harry has in the sea.
 b. Mary has fished in the ocean and I think Harry has in the sea.

In Berwick and Weinberg (1984), we claimed that bounding was expected in gapping constructions because the complements of the gapped verb had to be correctly attached in VP internal or external position. Correct attachment depends on properties of the verb. Since an overt verb is not available to direct the parser in a gapped constituent we predicted that deterministic attachment of these complements required a look at left context (some previous conjunct containing an overt verb). Given the usual requirement that we assume only bounded access to this left context, the bounding constraint on these constructions followed. Since the parser faces the same problem in both types of gapping constructions, Fodor (1985) correctly argued that our own assumptions would predict that bounding constraints apply to gapping in conjuncts with or without auxiliaries since both structures

contain potentially ambiguous complements in gapping constructions. In addition, we claimed that the parser could only have access to c-commanding left context and since the PP complement in the first conjunct of a gapped structure does not c-command material in the second conjunct, it could not guide attachment in this clause. Therefore, we must show that complement attachment of PPs does not require access to left context but that there are other properties of gapping constructions that require this access only in cases where no overt auxiliary precedes the gapping site. Before turning to this point, let us present some additional evidence for our assumption that the parser can make a decision to create or attach an empty category on the basis of c-commanding information alone.

Consider a case like (41a) with the structure (41b):

- (41) a. John begged ELOISE^{NP} to get married and FRANCES to get divorced.
- b. John [_{CP} [_{CP} begged_i [_{NP} ELOISE] [_{PRO} to get married]] and [_{CP} O_i [_{NP} FRANCES] [_{PRO} to get divorced]]]

We assume that lexical information about the head of a category is projected from that head to its most "maximal

projection".²³ Therefore, information about the verb is projected to the VP dominating both conjoined clauses and is available to guide parsing decisions about the structure of the second conjunct. In particular, the fact that 'beg' is a verb that takes an NP S complement is available and we predict that the phonetically present remnants 'Frances to get married' will be placed in an appropriate syntactic structure, thus facilitating interpretation. Notice however, that we can also conjoin full IPs which include gapped verbs in the second conjunct. In these cases, we predict that since only information about the inflection, the head of the sentential category, will percolate by standard conventions, information about the verb and its complement structure is not available as a cue about the structure of the second conjunct. Thus in a case like (42), the fragmentary structure (43) will be created.

(42) JOHN begged ELOISE to get married and HAROLD SUE

(43) [_S [_S John [I [begged Eloise to get married]]] and [_S Harold [I [_{VP} [O] [_{NP} Sue]]]]]

We must assume that the correct interpretation of this structure, which includes the final complement of the verb 'beg' as part of the second conjunct applies at a subsequent level of representation.²⁴

If this is correct, then we would predict that there should be cases where the parser will not have the appropriate information in the second conjunct to allow it to create the correct syntactic structure. This prediction is confirmed by the contrast between (42) and (44).

(44) *John begged ELOISE to get married and HAROLD SUE to get married.

In (44) the parser does not have access to the lexical structure of the verb of the first conjunct when it reaches the second conjunct. Given the minimal commitment theory, it has no evidence to create a separate NP and S and would build the structure (45a) instead of (45b).

(45) ... and [_SHAROLD [_{VP}0 [_SSUE to get married]]]
 [_SHAROLD [_{VP}0 [_{NP}SUE] [_SPRO to get married]]]

This structure is correct for a case like (46) and predicts the relative acceptability of this case.

(46) John [believed]₁ SUE to be funny and HAROLD [O]₁ MARY to be smart.



The Libraries
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Institute Archives and Special Collections
Room 14N-118
(617) 253-5688

This is the most complete text of the
thesis available. The following page(s)
were not included in the copy of the
thesis deposited in the Institute Archives
by the author:

Page 105
106
107

be the only permissible expansion of the embedded clause in the second conjunct. Since 'believe' doesn't take small clauses we predict the unacceptability of the structure in contrast with the acceptable (54b).

- (54) a. *I think Fred is a fool and Sue believes John
stupid.
- b. I think Fred is a fool and Sue believes John is
stupid.

In contrast, cases that involve only main verb deletion will never create the same kind of ambiguous situations. This is because the presence of an overt auxiliary unambiguously signals that a verb phrase must follow. One never finds overt auxiliaries in small clauses. Since the parser will always be correct if it expands the phrase after an overt auxiliary as an empty headed VP, it will never have to scan the left conjunct. In a case like (55) it simply uses the presence of the locally available overt auxiliary to decide about subsequent expansion of the tree.

- (55) John has fished in the ocean and Bill has [_{VP} [_{VP}]
in the sea]

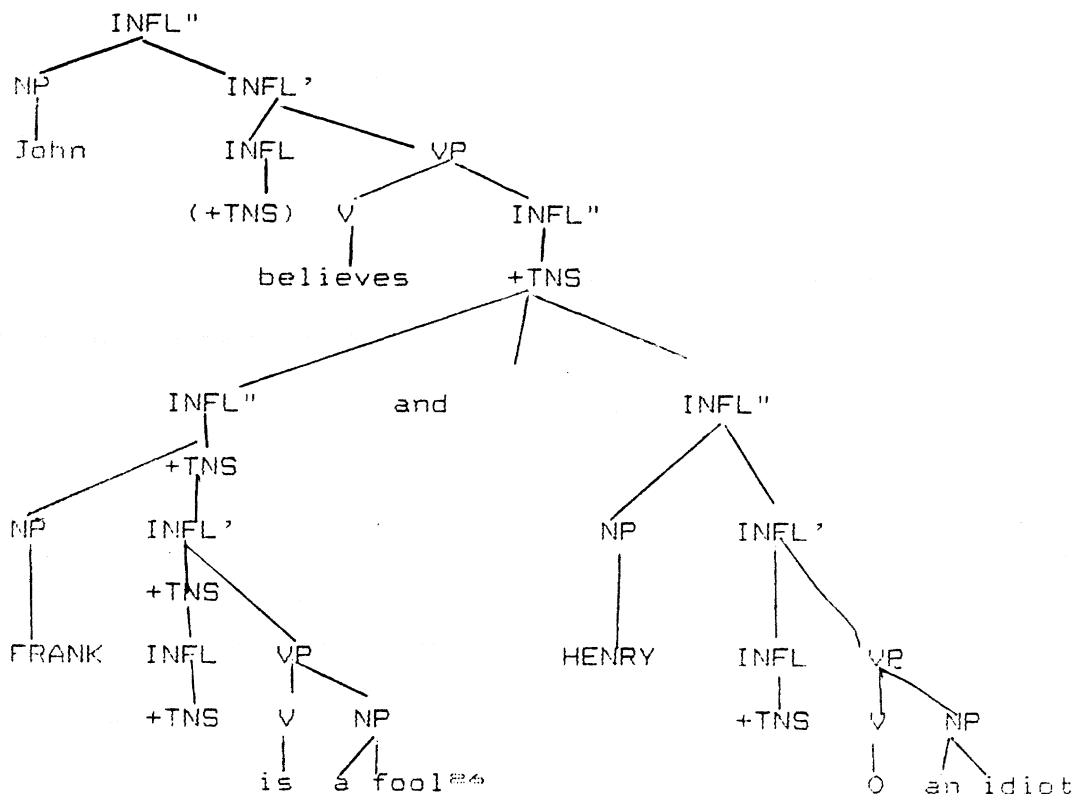
Since we never need to scan left context when the auxiliary remains in the surface string, we do not expect Main Verb

Deletion to obey bounding constraints. This is in fact what Fodor observes.

This account has another virtue in that the information provided by the left context to resolve the ambiguous cases will be available at the time when the parser is confronted with the ambiguous material of the second conjunct. This contrasts with the analysis of Berwick and Weinberg (1984) where, as we mentioned before and as Fodor (1985) correctly notes, correct identification of a verb's subcategorisation and selectional properties involves scanning the actual verb of the previous conjunct. Unfortunately our parser will not have access to this material. As mentioned above, our processor can only consider c-commanding information as the basis for future decisions about tree expansion. This is no problem for the present analysis because we distinguish small clauses from gapped constituents merely by scanning previous conjuncts for the presence of a tensed auxiliary. If we treat sentences as the maximal projections of INFLECTION, as suggested in Chomsky (1981) and if we assume that lexical information about the head of a category is projected from that head to its most "maximal projection", then the relevant information will percolate up to the highest S node on the tree and thus be available to the parser for tree expansion decisions.

Consider again (52), repeated as (56) with irrelevant details omitted:

(56)



By the time the parser reaches the locally ambiguous second conjunct, the first conjunct will have been completed and shunted to the second storage representation. Thus information contained in this conjunct will not be available for decisions about tree expansion. This causes no trouble though because we see that the fact that the inflection of the first conjunct is tensed can be read off the highest INFL'' which c-commands and is boundedly far away from the INFL' of the next conjunct. If the first conjunct was a

small clause, then the O-inflection would also percolate up to the maximal S node. This is all the information that the parser needs to correctly expand the tree of the second conjunct. If the previous conjunct contains a tensed or infinitival inflection, the parser expands the second conjunct as a gapped structure. If the previous conjunct contains a O inflection then the parser expands the ambiguous structure as a small clause. The analysis makes the interesting prediction that if S's instead of Ss are conjoined, tense deletion should be unacceptable. Since S' is not a projection of Inflection, conjunction of S's would not allow percolation of information about the inflection beyond the first conjunct in a structure like (56).²⁷ Since expansion as a tensed structure is conditioned by the presence of an overt auxiliary in the previous conjunct, the parser will not be able to apply the rule. This prediction is confirmed by comparing (57a-b), where we have conjoined Ss, with (57c-d), where we have conjoined S's.

- (57) a. That Bill would hit Sam and Frank would hit Harry surprised me.
- b. That [_SBill would hit Sam] and [_SFrank [_{INFL}-O [_{VP}[_{VO}] Harry]] surprised me.
- c. That Frank would hit Sam and that Bill would hit Harry surprised me.

- d. * $[_S [_S \text{That} [_S \text{Frank would hit Sam}]] \text{ and } [_S \text{that } [_S \text{Bill } [_{INFL} \text{O } [_{VP} [_{OO} \text{Harry}]]]]] \text{ surprised me.}$

As predicted, Main verb deletion can apply in both conjoined S and S's as shown in (58):

- (58) a. That Frank would hit Sam and Bill would Harry surprised me.

- b. $[_S \text{That } [_S \text{Frank } [_{VP} \text{would hit Sam}]] \text{ and } [_S \text{that } [_S [_S \text{Bill } [_{I} \text{would } [_{VP} \text{O Harry}]]]] \text{ surprised me.}$

Thus this approach correctly distinguishes the two cases of gapping.

3.8.3 Complement and Adjunct Attachment

Returning to our first problem, we must show why complement vs. adjunct attachment which applies in both types of gapping does not force the parser to look at left context, thus incorrectly predicting that bounding constraints apply to both types of gapping. The treatment in Berwick and Weinberg (1984) assumed that the semantic interpretation of adjuncts and complements proceeded in essentially the same way, by reading off tree structure. If we assume this, then it follows that a deterministic parser must attach PPs and other

adjunct phrases as they are attached in the grammar, in order to carry out semantic interpretation. However, this assumption is highly dubious. As Miller and Chomsky (1963), Marcus (1980) and many others note, in certain cases, strings of PPs can occur in potentially an unlimited variety of configurations. Thus a sequence like (59a) can have any of the interpretations in (59b-d).

- (59) a. The man in the house by the river in the woods near
the town
- b. in the house[by the river [by the woods]] [near
the town]
- c. in the house [by the river [by the woods [near the
town]]]
- d. [in the house[by the river]] [by the woods [near the
town]]

A parser that had to do semantic interpretation from tree structure would find itself in an exponential regress in cases like these. In order to figure out which interpretation to give the sentence, it would have to compute the correct syntactic structure, but in order to do this it has to compute all the possible patterns compatible with the string, and then see which one "it means to say". This will

cause a corresponding exponential slowdown in the parsing algorithm, if all trees must be explicitly reconstructed. To solve this problem, these authors suggest that adjunct phrases that can be ambiguous (either between adjunct and complement readings or between various adjunct readings) should be parsed as essentially flat structures. Semantic subroutines can then come in later and decide between possible readings: a procedure that allows us to maintain efficient parsing.

Put in the context of gapping constructions, if a parser cannot figure out where an adjunct is attached from local context, it can simply attach it as a flat structure to the lowest node in the parse tree. Then independently needed semantic routines will give this phrase its appropriate semantic interpretation. Given this story, complement phrases that cannot locally be distinguished from adjuncts do not cause the parser to use left context to disambiguate their interpretation. The independently needed semantic subroutines do that. Therefore attachment of potential adjuncts does not imply boundedness of a construction, thus correctly allowing us to maintain the previously motivated distinction between Tense and Main Verb Deletion.

3.8.4 Subjacency as an S-structure Condition

We have assumed that the Subjacency condition is motivated by processing considerations. Moreover, we have assumed that this condition is part of universal grammar. This assumption is crucial because we have claimed that the formulation of the bounding condition imposed by the parser is governed by restrictions imposed by universal grammar on its rule writing formalism.²² If Subjacency were not part of UG then there would be no reason for why restrictions imposed by UG should apply to this condition.

Given these assumptions, we must insure that we can state the Subjacency condition in a way that allows it to apply correctly to the gapping cases discussed immediately above. In particular, we must insure that the parsing instantiation of the Subjacency condition does not need to use illicit rule writing formalism to distinguish the application of Subjacency in the case of tense deletion from its non-application in verb gapping. We do this in three steps. We show that the verb deletion part of the gapping rule does not result in the formation of a connection between the gapped verb and its controller. Next, we show that tense deletion requires such a connection between a present and gapped INFL. Thirdly, we assume with Chomsky (1986) that Subjacency applies to S-structure chains. The effect of these assumptions will be the enforcement of the Subjacency condition in cases where the tense is deleted (where no

auxilliary is present) but a lack of bounding effects in cases where the auxilliary remains.

Let us assume the formulation of subjacency provided by Chomsky (1973) as condition on UG:

(60) No rule can relate X and Y in the structure

[.... Y [... [... X] ...] Y]

where and are bounding nodes.

Recoded as a parsing rule this condition reads:

(61) If [Y] a cue for the creation of a category X appears in a local left context, then create X.

Following Chomsky (1986), we assume tht Subjacency is a condition on S-structure chains. As discussed in chapter 2, and as we will discuss below, a c-command condition is incorporated into the Subjacency Constraint.

Given our assumptions about gapping constructions and the formation of conjoined structures, a verb in a conjoined IP structure does not c-command another verb in any other conjunct. Thus the creation of a verb by the parser in a gapped structure cannot depend on the presence of a verb in another conjunct. At the formal level, the gapped verb cannot form a chain with a verb in another conjunct because

the c-command condition would not be respected by links in the two conjuncts. The only left context that the parser can refer to in creating a gapped VP or other predicative phrase is the inflection (or lack of inflection) that appears within the same conjunct as the gapped predicate. This information will always be locally available to the parser.

Things are different in the case of the creation of the INFL marker in a conjoined structure. As mentioned above, information about the structure of the inflection marker of any conjunct in a conjoined IP structure is available in all conjuncts at s-structure because this information can be percolated to the maximal projection that c-commands all conjuncts in a conjoined IP structure. Thus the value of the INFL marker ($\{\pm \text{tense}\}$ or 0) can depend on the value of the INFL in another conjunct as that information is available at s-structure. Moreover, the only way to tell the value of the inflection marker is to look at another conjunct because a conjunct with a gapped INFL marker has no information about the INFL's value. Since this information involves coindexing (and chain formation) between INFL markers, we predict correctly that this operation will be governed by Subjacency.

3.8.5 Subjacency as a Natural Condition on Grammars

Even though we have argued that we can state the Subjacency condition as a condition on S-structure chains using natural predicates of grammar in a way allows it to apply correctly to the cases discussed so far, we must still appeal to principles external to grammatical theory to explain why certain constructions must form chains at S-structure.

Thus for example, we assume that the coindexation associated with chain formation occurs automatically in the case of S-structure movement, so there are formal reasons why Subjacency should apply in these cases. Notice though that we cannot say that subjacency is a condition on CHAIN formation in general because, as we showed above, this condition does not apply to the chains formed by LF movement. It also does not apply to the chains created by the Binding Theory (arguably at LF).

Moreover, in the case of parasitic gaps, the chain formed between the parasitic gap chain and the real gap chain is not an automatic consequence of movement. Therefore, although the grammar can represent the fact that a composed chain is

formed at S-structure, it cannot explain why this chain must be formed at S-structure.

Our parsing explanation solves these puzzles. The elements in a chain formed between a lexical anaphor and an antecedent are independently detectable without recourse to left context. Therefore we predict that these elements would not need to be incorporated under the formal definition of subjacency and thus there would not need to be any principle forcing the construction of these chains at S-structure.

Similarly, the empty operator in a parasitic gap structure cannot be unambiguously created without recourse to left context. Therefore, we would predict that an LR(k) parser would need to incorporate conditions forcing the creation of a composed chain that would be governed by subjacency at S-structure. The parsing explanation of the Subjacency constraint thus can be seen as applying to UG and as acting to force UG to designate the class of constructions that obligatorily form chains at S-structure.

3.9 Bounding Conditions in LF

In the section on Gapping we have argued that Subjacency cannot be seen as a natural condition on movement and only

has a natural characterisation in terms of the problems these rules create for the natural language parser. We argued that this problem was created only if the construction under consideration entailed a parsing decision about tree expansion, thus correctly predicting that Subjacency should not apply to lexically specified elements, pronouns, anaphors, and constructions involving WH in situ. A direct challenge to both of these claims is posed in an interesting paper by David Pesetsky. Pesetsky (1985) argues that Subjacency is a defining characteristic of movement and as such applies to LF Movement as well. Our analysis of gapping shows that bounding is not a defining characteristic of movement, thus removing the conceptual motivation of Pesetsky's claim. If LF movement can be shown to be governed by Subjacency, we will be in the unhappy position of having no natural characterisation of the constraint. LF Movement does not involve tree expansion (the category moved is "in situ" on the surface) and so we do not expect the constraint to apply in these cases. In fact, there are many superficial violations of the constraint in LF. However, Pesetsky reanalyses these cases in a very interesting way; claiming that these superficial violations are not violations at all because they are not derived by Movement. Pesetsky claims that the domain of an LF item is specified either by movement or a process of linking to a Q operator.³⁰ A case like (62)

below, that violates Subjacency is thus to be derived by linking.

- (62) a. Which man believes that [_S[_{NP}pictures of which women]
are on sale]
- b. [_{COMP}Q_{i,j} which man_i [_S_e_i believes that [_S
[pictures of [which women]_j] are on sale]]]]

The scope of the second quantifier is marked by its being linked to the Q marker in the matrix complementizer and so does not involve movement. Pesetsky predicts that all quantifiers whose scope is derived by linking without movement should allow subjacency violations. To show that Subjacency holds in LF, Pesetsky then has to present a class of quantifiers whose scope can only be derived by movement because he only predicts that this class of quantifiers should show overt Subjacency effects. Using a set of independent criteria, Pesetsky argues that only "discourse-linked" ("d-linked") WH words can use the linking mechanism to represent their scope. A discourse-linked element is one whose reference must be taken from the set of elements mentioned in a prior discourse. Since it has referential properties it is really not a quantifier at all, and therefore we might not expect it to undergo the usual quantificational process of scope assignment. Non-discourse

linked quantifiers do not have these referential properties and do not have to refer to an element mentioned in a prior discourse. We therefore expect them to use the normal means of scope assignment, which is movement.³¹ A typical d-linked quantifier was given in (62) above. An answer to a question of the form "which NP" is bizarre if it does not pick from a set of previously mentioned NPs. So for example (63b) is a bizarre answer to (63a) if the men previously mentioned were Sam, Frank, and Tim.

- (63) a. Speaker 1: Which man do you like?
 b. Speaker 2: Bob

If we use a non d-linked quantifier, the same answer is not bizarre even if preceded by the same discourse.

- (64) Who do you like? Bob.

Thus the prediction is that we should see clear Subjacency effects if we choose non d-linked quantifiers. Pesetsky constructs cases using the quantifier "what the hell", a quantifier that clearly does not ask for a discourse internal reference, and then he shows that indeed locality constraints do show up. The examples are given in Japanese, a language that, as shown by Huang (1982) and Lasnik and Saito (1984) has WH movement only at LF and can "violate" Subjacency when

other quantifiers beside "what the hell" are involved. The relevant examples are given in (65).³²

- (65) a. Mary-wa [_{NF} [_S John ni nani-o ageta] hito-ni]
 MARY-TOP JOHN-DAT what gave man-DAT
 atta-no?
 met-Q?
 What did Mary meet [_{NF} the man [_S who gave e_i to John]
- b. Mary-wa [_{NF} [_S John -ni ittai-nani-o ageta]
 Mary-TOP John-DAT what the hell gave
 hito-ni] atta-no?
 man-DAT met-Q?
 What the hell did Mary meet the man who gave to
 John?
- c. Mary-wa [John-ga nani-o yomu mae-ni] dekekata -no?
 Mary-NOM John-NOM what-ACC read before left-Q?
 What_i did Mary leave before John read e_i
- d. Mary-wa [John-ga ittai-nani-o yomu mae-ni
 Mary-NOM John-NOM what the hell read before
 dekekata-no]?
 left-Q?
 What the hell did Mary leave before John read?

While the locality effects in these examples are clear, it is not so clear that they should be accounted for in terms of Subjacency. In French, one can test whether binding constraints result from Subjacency or some other locality condition by seeing whether the construction respects the WH-island constraint of Ross (1967). As Rizzi (1982) and Sportiche (1981) have noted, French and Italian allow violations of the WH Island Constraint as shown by example (66a). (66b) shows that the locality conditions on "what the hell" phrases in these languages are stricter, suggesting that a different locality condition must govern their distribution.

(66) a. Qu'est-ce que_i tu as dit comment [manger e_i]

What_i you said how to eat e_i?

What did you say how to eat?

b. *[Que diable]_i as-tu dit comment [manger e_i]

[What the devil]_i said you how to eat e_i?

What the devil did you say how to eat?

Aoun (personal communication) presents some data that suggests that these examples violate the locality restrictions imposed by his "Generalised Binding Theory"²². We will discuss this theory in great detail below. Aoun

(forth) generalises the binding theory to include both A and A' anaphors. He explains Huang (1982)'s insight that subjects and some adjuncts must both be bound in a local domain by claiming that both are A' anaphors and so must be bound in their minimal domain. In (66) the minimal governing category is the embedded S. In Aoun's theory, an operator or trace in COMP may act as an A' Binder. Since the lower COMP is filled with another complementizer and since French doesn't allow doubly filled COMPs, the phrase "que diable" would have to move in one swoop to the matrix COMP. Thus the trace of "que diable" is not bound in its minimal governing category and the sentence is correctly ruled out. Interestingly, it is crucial to Aoun's analysis that phrases like "que diable" not be referential in a way reminiscent of Pesetsky's claim that such phrases are non d-linked. Aoun must distinguish objects from adjuncts and subjects because objects are not governed by the same locality conditions. He does this by claiming that objects do not have a domain because they don't have an accessible subject. If agreement were the accessible subject in a case like (66), then we should be able to coindex the object with it. However, by convention, the subject is also coindexed with agreement and so the A' anaphor would also be coindexed with the subject. In this case though it would be A bound. These A' anaphors are subject to Principle C of the Binding theory though and so cannot be A-bound. These A' anaphors can simultaneously

satisfy Principles A and C of the binding theory, only if they have no governing category. Superficially the same reasoning should apply to phrases like "what the hell" phrases and other adjuncts. However, Aoun has pointed out that only "referential expressions" are subject to Principle C of the Binding Theory. If a phrase is not referential it does not obey principle C and so it does have an accessible subject and must be bound in its local domain.³⁴ Thus it seems that Pesetsky's distinction between "referential" (d-linked) and non-referential (non-d linked) phrases is in fact a theoretically relevant distinction, but one that should map into the theory of Binding and not Subjacency. Subjacency is irrelevant to explaining the distribution of "what the hell" phrases.

Pesetsky's analysis must also deal with further complications because many d-linked phrases can also be used in non d linked contexts. In Japanese, many WH phrases that don't seem to be linked to a discourse element also seem to move out of an island. Pesetsky notes though that there are subtle facts that suggest that Japanese speakers actually obey Subjacency and get the "problematic" readings by pied piping the entire island along in LF. This explains why speakers can answer a superficial island violation like (67a) with (67b) but not with (67c).

(67) a. Mary-wa [John-ga nani-o yomu mae-ni] dekaketa-no

What did Mary leave before John read

b. sensoo to Heiwa-o yomu mae desu

War and Peace-ACC read before COP

It's before he read War and Peace

c. Sensoo to Heiwa desu

It's War and Peace

With or without Pied-piping Pesetsky's analysis has problems with English. Consider (68).

(68) Who wonders whether the sheik bought what?

We can give a wide scope reading to "what" in this sentence even though movement of this phrase would violate the WH-island constraint which English respects as shown by the unacceptability of (69).

(69) What does Mary wonder whether John saw?

Moreover, this sentence can be used in a non d-linked context as the following discourse shows.

Speaker #1: The sheik returned from a big trip to London.
 I think everyone is wondering what he did
 there.

Speaker #2: You mean "who wonders whether the sheik
 bought what".

Moreover, the "pied-piping" analysis runs into problems vis a vis the ECP. As noted by Aoun, Hornstein, Sportiche (1981) a case like (70) is ungrammatical because it violates the ECP at LF.

(70) *Who said that what was exhibited. ~~??~~

A sentence like (71) on the non d-linked reading will violate subjacency if the WH phrase in situ moves alone to COMP in LF. If we pied-pipe the whole subject NP though, we predict that this sentence should also violate the ECP. There is no way to derive the grammatical reading, assuming Pesetsky's analysis.

(71) Who said that pictures of what were exhibited

We have shown that the locality effects for "what the hell" type phrases follow from the binding theory, thus removing the need to invoke subjacency at LF for these cases. We have also shown that the pied piping analysis needed to make the account work for a variety of cases makes wrong predictions in English. We conclude that we should not extend subjacency to the LF domain and that the condition maintains its natural parsing interpretation.

3.10 Conclusion

In this chapter, we argued that an explanation both for the existence of the Subjacency constraint and the class of cases that this constraint applied to could be derived from functional constraints of efficient parsability placed by the use of bounded context on deterministic parsers. We showed that the class of cases where bounding restrictions applied was unnatural from a grammatical perspective, but fully natural given the perspective of this chapter. Finally, we provided an alternative to Pesetsky (1985)'s analysis of LF movement which was consistent with our claim that since LF movement does not lead to ambiguous structures, this type of movement should not be governed by bounding theory.

Footnotes

1. This topic will be taken up in more detail in Chapter 7.
2. This discussion is a condensation of material from Berwick and Weinberg (1984) pp.183-192. and Berwick (1985) pp.313-332.
3. An LR(k) grammar actually instantiates a typical "bottom up" approach to tree expansion in that the grammar starts with the rightmost terminal symbol in a string and expands it to its most immediately dominating non-terminal category. This process continues and results in a successful derivation if all terminal nodes can be associated with nonterminals by the set of stored phrase structure rules and the final association (called a reduction) derives a matrix sentential nonterminal. An example of this sort of derivation is given in (a)

- (a) 0. Start \rightarrow S
 1. S \rightarrow NP VP
 2. VP \rightarrow V NP
 3. NP \rightarrow Mary, John
 4. V \rightarrow loved

We can use these rules to produce a rightmost derivation for:

5. Mary loved John

The first move reduces the rightmost terminal symbol according to the stored rules to an NP, leaving the string
 Mary loved NP

The next terminal can be reduced to a Verb by rule a4 above yielding the string

Mary V NP

The string V NP can be reduced to a VP by rule a2. The final terminal 'Mary' can be reduced to an NP by a3, and the string

NP VP can be reduced by a1. the output of a1 can be reduced to the start state (which is also a final state) and so the derivation is successful and the string is generated. The (R) in the name LR(k) stands for the fact that this parser does not predict the presence of terminal items from access to the nonterminal symbols available in its stored grammar rules. Rather, it accesses nonterminal material directly and tries to associate this material with a unique derivation from the set of possible sets of stored phrase structure rules. The parsing analogue of this grammar mimics the property of direct access to terminal material but reverses the order of the computation, trying to reduce the leftmost terminal symbol first, instead of the rightmost.

Thus it computes a rightmost derivation in reverse, proceeding in a left to right fashion.

4. Aho and Ullman (1972) define a grammar as LR(k) if "...when examining a parse tree for G" (the grammar(ASW)) "we know which production is used at any interior node after seeing the frontier to the left of that node, what is derived from that node, and the next k terminal symbols.

Aho and Ullman pg. 379

5. See Lewis and Papadimitriou (1981) or any standard formal language theory text for an explanation of why the use of essential variables extends the expressive power of the system beyond the class of regular expressions that a finite state language can generate.

6. A simple parallel approach would predict that there would be no difference in acceptability between cases like (a) and (b), which, as we discussed in the previous chapter is incorrect.

(a) Who_i did the little girl beg e_i [PRO_i to sing those stupid French songs] (as a first interpretation for this structure)

(b) Who_j did the little girl_i beg [PRO_i to sing those stupid French songs for e_j]?

7. Berwick and Weinberg (1984) state the subjacency Constraint as follows:

If adjacent (S/S') contains a WH element or trace, expand an empty NP position as a trace.

The standard Complex NP and Wh-islands fall out under this account because the parser will not find the category in the adjacent COMP that will trigger placement of a trace in the ambiguous position. The adjacent COMP in a complex NP will be filled by an operator predicated of the head of the NP as shown in (a) and in the WH-island construction, by a WH bound to another open position as shown in (b).

a. Who_i did [you [_{OP} believe [_{NP} the fact_j [_{CP} that [John liked e_i]]].

b. Which man_i did John wonder [which woman_j [e_j loved e_i]].

The well known parameterisation of the bounding nodes for this constraint can be stated exactly as they are in previous accounts.

8. We will formalise this condition and indicate an explicit theory of bounding nodes below.

9. The PRO in (c) is locally detectable because, given the Projection Principle, the embedded phrase 'helping his country' requires a subject and given case theory the empty category placed in this position must not need case. Given theta theory, the empty category cannot be an NP trace because the subject position is theta marked. The only possible choice is PRO.

10. See Fodor(1985) for a detailed discussion of this point.

11. Here we differ from Chomsky (1986) who licenses empty operators if they are subjacent to the real gap as opposed to the real operator. We will justify this assumption below.

12. This analysis is refined in Chapters (5) and (7) below.

13. This assumption was also adopted in the previous analyses of Engdahl (1981) and Chomsky (1982).

14. The reason for this is the main focus of Chapter 7.

15. See Chomsky (1982), Evans (1980), Higginbotham (1980), Hornstein (1984) and Haik (1985) for various analyses of the "pronoun as bound variable" phenomenon.

16. We will see in Chapter 6 why (a) is equally unacceptable:

(a) *Every contra_i thinks that the bastard's_i mother should get an M-15.

17. (30) may sound slightly worse than (29a or 29b). This is because as argued by Koopman and Sportiche (1982) this sentence violates the Bijection Principle. The crucial point is that (30) should sound no worse than (a).

(a) Mary buttonholed every senator_i before he_i could get away.

As Koopman and Sportiche argue, Bijection violations result in weak unacceptability.

18. Recall that (b) and (c) will not be perfect as they are Bijection violations. The crucial point is that they are not worse than the corresponding sentences where they are replaced by pronouns.

19. Richard Larson (personal communication) suggests that cases like the following are disconfirm the claim that the grammaticality of cases like (30a and 30b) is related to the c-command properties of the real gap in the VP.

(a) No wine was sold before the damn thing was ready to drink.

(b) No wine soured before the damn thing was brought to market.

Possible LF structures for these cases are (c) and (d). The cases are problems because the variable seems to be in a c-commanding position but nonetheless the cases are acceptable in apparent violation of Principle C and suggesting that something intrinsic to the adjunct structure protects the epithet from such violations.

(c) No wine_i [_{TP} x_i was sold [_S before the damn thing_i was ready to drink]].

(d) No wine_i [_{TP} x_i soured [_S before the damn thing_i was brought e_i to market.]]

First, notice however that cases like (e) and (f) are significantly worse and so adjuncts can't always block Principle C violations.

(e) *No spy_i escaped before we located the bastard_i.
 *[Every bottle of milk]_i glistened after we polished the damn thing_i.

vs.

(f) No spy_i escaped after we located him_i.
 [Every bottle of milk]_i glistened after we polished it_i.

Two factors seem to combine to make (c) and (d) as acceptable as they are. Notice first that in the quantified expression gets its theta role from the post verbal position in both of these cases. We might claim that these cases are interpreted with the variable in the postverbal (and thus non c-commanding position). In effect, this extends the treatment of quantifier lowering proposed in May(1985) to these cases. Thus the structure of (c) and (d) would be (g) and (h)

(g) No wine_i [_{TP} Pro [_{VP} soured e_i] [_S before the damn thing_i [_{VP} was sold e_i]]
 was ready to drink.

As noticed in LGB (pg. 177), quantifier lowering cannot take place if the quantified expression appears in a theta position at s-structure. Thus we would predict that we should not be able to treat an epithet as a bound variable if it is bound to a quantified expression that appears in a c-commanding theta position at s-structure. Lowering should

be possible only from a non-theta position. This prediction seems to be confirmed by the contrast in (h) and (i) which have the LF representations (j) and (k).

(h) * No spy_i expected to be tortured after the bastard_i confessed.

vs.

No spy_i expected to be tortured after he_i confessed.

(i) No spy_i was likely to be tortured after the bastard_i confessed.

No spy was likely to be tortured after he_i confessed.

(j) No x, x a spy: x_i expected [PRO_i to be [tortured e_i] after the bastard_i confessed.

(k) No x, x a spy: pro_i was likely [to be tortured x_i] after the bastard_i confessed.

This analysis predicts that cases like (f) where the quantified expressions are not subject to lowering should be worse.

Another possible confound comes from the fact that 'No wine' in (c) and (d) can be given a mass or count interpretation. It is unclear that coreference to a mass noun is an instance of binding as shown by the relative acceptability of (l) over (m):

(l) Humankind_i threatened humankind_i during WWII.

(m) *The soldier_i threatened the soldier_i during WWII.

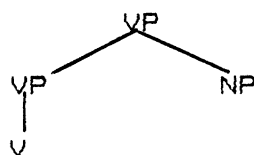
The second example in (f) where the epithet clearly refers to 'the bottle', a count noun, which is the subject of an intransitive verb and is thus not subject to lowering takes both of controls for both of the potential confounding factors cited above and seems not to permit the bound variable interpretation for the epithet. When these factors are taken into account, it seems that we can maintain the claim that the bound pronoun rule must apply to the LF representation and that the anti-command requirement between R-expressions in A-positions, is uniformly respected at this level of representation. The bound epithet can be bound by the quantified expression that appears in an A' position at LF, but the variable itself must not c-command at this level of representation.

20. Contreras also argues that parasitic gaps should be bound to empty operators. We have adopted this part of his analysis.

21. Contreras 1985 #4.

22. We follow the convention of indicating the placement of heavy stress on a word by capitalizing it.

23. Projection to the most maximal projection is supported by the fact that movement of a postverbal subject in Italian is permitted. Since these elements occur in structures like (a) we must insure that the verb can transmit its features to the maximal VP in order for the trace of the post verbal subject to satisfy the conditions on proper government imposed by the ECP.



24. That is, we assume that as long as the syntactic component outputs a complete constituent, some missing elements of the complement structure can be added at the post s-structure level. We might assume with Williams (1978) that missing portions of a conjoined structure can be inferred from the complete conjunct, as long as the remains of the the incomplete and complete conjunct can be given an identical factorization. In any case, regardless of the mechanism we choose to implement this idea, there are many reasons to treat these complements as a special phenomenon. First, as Stillings(1975) notes, this type of deletion seems to be limited to complements in constituent final position. Thus (a) is much better than (b)

(a) John dreamed that SHIRLEY was killed [in HIS sleep] and HARRY that SUE was in an accident O_i.

*(b) John dreamed [that SHIRLEY was killed]_i in HIS sleep and HARRY O_i in his daydreams.

Secondly, these complements can be interpreted even if the head of the phrase that they appear in is not gapped. Normally complements can only be deleted along with their heads.

(c) I begged FRANK to get married, and Myron begged JAMAL.

Finally, the complement can be indefinitely far away from the "controller" in the previous conjunct:

(d) I begged FRANK to get married and I think that Sue said that MYRON begged JAMAL.

We will see below that normal gapping cannot occur in these circumstances.

25. The structures of small clauses is the subject of some controversy. Chomsky (1981) following Stowell (1981) argued that the embedded categories "Bill a fool" formed a sentential complement (in this case, with the structure $NP[NP\text{John}]_{NP}\text{ a fool}]$). Williams (1980) and (1983) argued that these categories did not form a constituent and that they were properly analysed as [... $[NP\text{ John}]$ $[NP\text{ a fool}]$...]. Hornstein and Lightfoot (1987) argued against Williams' analysis and in favor of a modified version of the Chomsky Stowell approach. The only point that is relevant to this argument however is that the predicates of small clauses are not VPs.

26. Chomsky (personal communication) suggests that the fact that locality violations in these constructions are relatively severe when compared to standard subjacency violations is evidence that the restriction on gapping is different from subjacency. It seems though that the severity of the gapping violations could come from the fact, in order to interpret standard subjacency violations one must simply add an empty category to a potential theta position. In order to insert a tense and gapped verb into a structure misanalysed as a small clause, we must turn a "0 inflection" into one that is marked as +TNS, we must also add a VP and reattach any other elements in the small clause as complements to this PP. Therefore the relative unacceptability of these structures may simply come from the more complicated nature of the "repair routines employed by the parser in these cases.

27. See Zubizarretta (1982) and Stowell (1981)

28. Namely, we have assumed that the grammar could only implement a bounding condition on left context as subjacency because the grammar does not allow counting predicates and we have assumed that subjacency governs all instances of WH movement because of a ban on Boolean conditions and existential quantification (see above).

29. Below, we will discuss reformulations of this condition proposed in Chomsky (1986). These reformulations do not change the logic of our remarks.

30. This mechanism was first discussed by Baker (1970), as Pesetsky notes.

31. Henceforth, we will borrow Pesetsky's terminology and refer to these elements as D-linked and non D-linked elements.

32. All examples in 22 are from Pesetsky 1985.

33. The Generalised Binding Theory is presented in Aoun (1981) and (1986).

34. See Wahl (1987) who crucially uses this argument to explain the distribution LF Movement of Adjunct phrases.

35. As Aoun, Hornstein, Sportiche (1981) argue, the WH element has to move here to get the appropriate "paired reading".

Two Versions of Locality

In the previous chapters, we have concentrated on the relationship between the grammar and the processor; arguing that processing considerations can influence the form of universal principles of grammar. Recent work, particularly by Chomsky (1986) has reconsidered the nature of bounding theory in an attempt to unify constraints on bounding with constraints imposed by the theory of government. As we will see, the version of subjacency proposed for example in Chomsky(1986) is inconsistent with the parsing theoretic view that we argued for in previous chapters; the view that this condition insures that there will be a bounded number of symbols between a trace and its local operator. In this chapter, we present the theoretical background that motivates Chomsky's analysis of bounding theory, critically evaluate that theory, and propose an alternative to it. In the next chapter, we will outline a theory of bounding that is consistent with the alternative proposed in this chapter.

The revisions to the theory of bounding found in Chomsky (1986) are part of an effort to unify the theories of government and bounding. Chomsky claims that there is a uniform set of categories that create domains into which

government is possible and out of which movement is allowed. The theory of government is relevant to the theories of Case and theta assignment, which apply under government and to the Empty Category Principle (ECP), which governs the distribution of non pronominal empty categories. Therefore, in order to evaluate the unification proposed, it is necessary to understand the theory of government that the bounding theory is supposed to tie into. In this section, I will review the theory of government, proposed in Chomsky(1986), concentrating on the parts of this theory devoted to proper government and the ECP. Next I will present the alternative to this theory that was originally proposed in Aoun, Hornstein, Lightfoot, and Weinberg(forthcoming)¹. Finally, I will outline Chomsky(1986)'s theory of bounding and some of the problems that the theory faces. More challenges to this theory will be taken up in chapter 6 after an alternative approach to bounding is discussed in chapter 5.

4.1 The Barriers Theory of Government and Proper Government

The Barriers approach modifies the traditional definition of the ECP given in (1).

(1) (a) properly governs (b) iff either

(a) governs (b) and (a) is a lexical category (that is, an X^0 in the X-bar system but not INFL, COMP or P)

or

(a) is a phrasal category X^n locally coindexed with (b).

Chomsky (1986) modifies this definition by making theta-marking a condition for lexical government and by modifying the domain that counts for "local coindexation" of antecedent government.²⁰ The previous definition assumed that S was not a maximal projection and defined local coindexation as coindexation between categories with no intervening maximal projections. The revised theory treats the projection of INFL and the projection of the complementiser as separate maximal projections (IP and CP) thus generalising X' theory to non lexical projections.²¹ Government and antecedent government can potentially be blocked by the intervention of any maximal projection if that maximal projection is a barrier. The definition of a barrier is given in terms of the notion of a blocking category. The relevant definitions are given in (2).

(2a) Blocking Category:

X is a blocking category for (b) if X doesn't have lexical sister that theta marks it and X contains b.⁴

(2b) Barrier:

X is a barrier for b iff either (1) or (2)

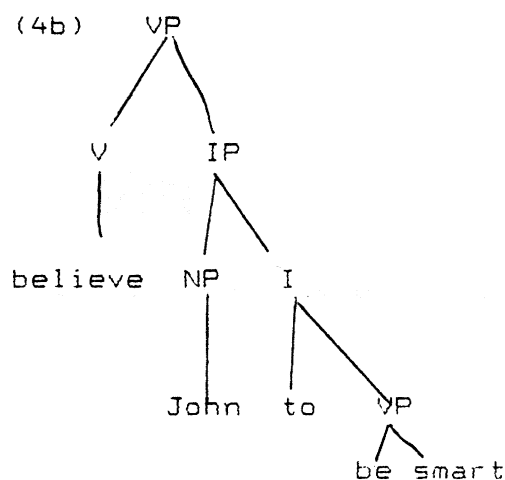
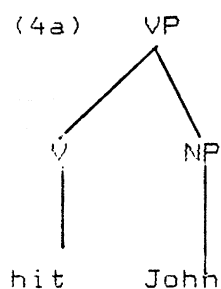
- (1) X immediately dominates Y, Y a blocking category for b.
- (2) X is a blocking category for b unless X is an IP.

The notion of government is in turn defined in terms of barriers as in (3)

- (3) (a) governs (b) iff (a) m-commands (b) and every barrier for (b) dominates (a)⁵

Given these definitions, a head will govern its complements and a lexical head will also be able to govern into a complement that it theta marks. Thus case assignment (which applies under government between a case assigner and assignee) is permissible within a phrase or in exceptional case assignment contexts. In the first instance this is possible because no maximal projections intervene between a head and its complements and in the second case the maximal

projection that intervenes between the case assigner and assignee is a complement of the assigning verb. Since it is theta marked by a lexical category it does not count as a blocking category for elements it contains. Relevant cases are given in (4) below.



Case assignment is not possible by a verb that cannot delete its complement CP because, even though IP cannot count as an inherent barrier, given part (b) of the definition above, it is a blocking category. Thus CP, which immediately dominates IP becomes a barrier through inheritance given part (a) of the definition of a barrier, even if it is not intrinsically a barrier, having been L-marked as the complement of V. Under this analysis, French, which has no CP deletion will not allow exceptional case marking into the subject position and therefore lexical subjects will not be permitted in infinitivals in French. A case in point is (5)

*(5) Tu crois[_{CP} [_{IP} Jean etre malade]]

As in previous approaches, a category can be properly governed if it is lexically governed or antecedent governed. One of the disjuncts of this condition must be satisfied at IF, which is the level where the ECP applies. Since proper government is taken to be a subspecies of government, these relations are defined in terms of barriers.⁶ A lexical category can lexically proper govern an empty position if it I-marks it, where I-marking assumes both that the empty position is not separated by any barriers from its proper governor and that the proper governor theta marks the empty category. As in previous analyses standard movement of an object to its complementiser position is allowed because the trace will be lexically governed by the verb that theta marks it.

Adding the notion of theta government to the notion of lexical government allows Chomsky to turn super-raising in NPs into an ECP violation.⁷

(6) John_i is certain that it seems [_{IP} e_i to be a fool.]

In this case "seems" governs the subject position of the infinitival clause. If theta marking is part of the notion

of lexical government though, this trace is not properly governed and the structure is correctly ruled out.

If we assume that lexical government involves theta marking, then adjuncts can never be lexically properly governed and movement will always be ruled out unless each link in the movement chain from adjunct position is antecedent governed. We will now turn our attention to the distribution of adjuncts and to antecedent government.

A category is antecedent governed if it is coindexed with an antecedent that governs it. In order to allow for movement from the subject position in a case like (7), Chomsky is forced to stipulate that IP cannot count as an inherent barrier because since the complementiser is not a lexical category it cannot L-mark this position and thus the presence of an IP would block antecedent government of the subject trace.

(7) [_{CP} Who_i [_{IP} e_i saw Bill.]]

WH- movement of an adjunct is still barred under this theory on the assumption that adjuncts hang from VP as argued by Andrews (1982) and others. In a structure like (8) the adjunct is not lexically governed by the verb because it is not theta marked by it. It is also not antecedent governed

because the intervening VP is not L-marked and thus counts as a barrier. The structure is given in (8).

(8) [_{CP} Why_i did [_{IP} Nory [_{VP} [_{VP} mow the grass] e_i]]].

In order to allow movement of a VP internal adjunct, and for other reasons to be discussed below, Chomsky allows maximal projections to adjoin to any X-max that is not an argument. In addition, he also allows a category to antecedent govern inside a maximal projection that it is adjoined to by redefining government in terms of exclusion. The definition of exclusion is given in (9a) and the redefined notion of government is given in (9b)

(9a) (a) excludes (b) if no segment of (a) dominates (b)

(9b) (a) governs (b) iff a m-commands (b) and there is no X, a barrier for (b) such that X excludes (a).¹⁹

Given these assumptions, the adjunct can first adjoin to the VP, which will no longer constitute a barrier for antecedent government. Its trace in turn will be antecedent governed by the trace in CP because IP is not an inherent barrier. We further assume that unless all parts of a category contain an element, that category does not contain the element and thus

cannot count as a barrier for an adjoined trace." The relevant structure is given in (10).

(10) [_{CP} why_i [_{IP} Nory [_{VP} e_i [_{VP} [_{VP} mow the grass]
e_i]]]]

The previous discussion has outlined how the ECP applies to the output of S-structure movement. Since the ECP applies at LF, we would expect both syntactic and LF movement to be subject to this condition. To show how this works, let us consider the case of "long movement"; movement of WH element across a clause with a WH in the intervening CP.

Long movement of an adjunct is barred under the assumptions outlined so far. An adjunct is by definition, not theta marked by a lexical category, and since antecedent government cannot take place over any barriers, long movement, either in the syntax, or in LF, will be barred due to a failure of antecedent government. Cases of long movement are given in (11a) and (11b) and the relevant structures, with the trace that fails to be antecedent governed underlined, are given in (11c).

*(11a) how did you wonder whether John fixed the sink.

*(11b) Who wonders whether John fixed the sink how.

(11c) [CP how_i [do] [IP you [VP e_i [wonder [CP whether [IP John [VP e_i [VP [VP fixed the sink] e_i]]]]]]

(11d) [CP how_i who_j [IP e_j [VP e_i [VP wonders [CP whether [IP John [VP e_i [VP [VP fixed the sink] e_i]]]]]]]] 10

In both cases the underlined trace fails to be antecedent governed because even though the IP is not an inherent barrier, it is a blocking category and thus can transmit barrierhood to the CP. Even though the CP is L-marked, it becomes a barrier by inheritance. Therefore, the next possible adjunction site (the matrix VP) is separated from the next trace in the chain by a CP barrier, blocking antecedent government and creating an ECP violation.¹¹

It should be noted that (11a) will be unacceptable in English because of considerations of bounding theory.

Chomsky (1986) correctly claims however that this violation is much more severe than long movement of an argument, which is also ruled out by subjacency, as we will see below.

Therefore, he claims that this case should be ruled out by another subtheory as well. The principles developed so far will also rule out long movement of an argument, which is an unfortunate result because such movement is possible in

many languages including some dialects of English. Let us consider a case where we have long moved a WH-object.

(12) what do you wonder who saw

As in (11) the trace at the foot of the chain will be antecedent governed by the trace adjoined to the most deeply embedded VP. Subsequent movement leaves a non-properly governed trace as shown in (13).

(13) [_{CP} what_i do [_{IP} you [_{VP} e_i [_{wonder} [_{CP} who_j [_{IP} e_j [_{VP} e_i [_{VP} saw e_i]]]]]]]]

In order to handle this, Chomsky follows Lasnik and Saito (1984) in claiming that traces of arguments in non-A positions can be deleted in LF. Therefore the trace that causes the ECP violation in (13) will be gone by the time antecedent government is checked. Again, following Lasnik and Saito (1984), Chomsky claims that traces of adjuncts cannot antecedent govern until LF. Therefore they must remain in the structure at this level and so long movement of adjuncts is ruled out at this level.

Although this proposal will work out technically there are various conceptual problems with it. Chomsky tries to

justify the stipulative difference about the level at which categories can be properly governed by reference to "the Projection Principle." One might think that it would be natural to force empty categories corresponding to arguments to be identified (perhaps by being properly governed) at all levels where they must be checked to see whether the Projection Principle is satisfied. This would force arguments to be properly governed at s-structure, but it is still not clear why these considerations disallow the mechanisms that properly govern adjunct traces from applying at this level. We will suggest below that the distribution of 'why' and 'how' follows instead from their semantic properties as "non-referential" categories, rather than from their classification as adjuncts. Independent evidence that the adjunct/ argument distinction is on the wrong track is provided in Wahl (1987) and Hornstein and Lightfoot (forthcoming).

Both works point out that all adjuncts do not seem to behave in the same way with respect to long movement. Thus (14) is ambiguous, with the WH-in situ capable of being construed with the WH element in either the matrix or embedded clause. This contrasts with (15), where the sentences are unacceptable.

(14a) Who asked what happened when

(14b) Who asked what happened where

(with the reading: 'Fred asked what happened at six o'clock, Sue asked what happened at seven o'clock...' or 'Fred asked what happened in San Diego, Sue asked what happened in Washington, ...')

*(15a) Who asked what happened why

*(15b) Who asked what happened how

Since temporals and locatives are not arguments and do not receive their theta role from the predicate¹², we would expect them to be able to be antecedent governed only at LF. But given Chomsky and Lasnik & Saito's assumptions about the presence of adjunct traces at LF, we would expect that long movement should be impossible for these cases as it is for the cases in (11).¹³

In this section, we have presented the bare bones of Chomsky's theory. We have shown how antecedent government and lexical government are assimilated under the theory of barriers. We have also presented empirical motivation for incorporating theta marking into the notion of lexical government by looking at super raising and adjuncts. In the next chapter, we will discuss why and how bounding theory is handled in this framework and look at further examples of movement in the syntax and in LF.

In the next section we will present an alternative account of the ECP that does away with some of the conceptually unsatisfactory aspects of the Barriers approach. A version of this account was first presented in Wahl (1987) and tries to handle the distribution of adjuncts in a different way. This account starts with an intuition that is similar to Chomsky's in that it tries to do away with an independent theory of antecedent government and tries to assimilate these phenomena to a subtheory that is needed independently in the grammar. It also tries to relate the theory of lexical government to the general theory of government. It is different from Chomsky's approach in that (i) antecedent government is assimilated to the theory of Generalized Binding as proposed by Aoun (1981), (1986) rather than to the theory of government. The disjunctive definition is dispensed with in favor of a theory in which both lexical government and Binding apply; the former in PF and the latter in LF.

4.2 The WAHL Framework:

A review of the cases discussed in the previous section shows that in the majority of cases, proper government can only be satisfied by antecedent government. In fact, in the final sections of Barriers, Chomsky dispenses with lexical government entirely, and assimilating all cases of proper

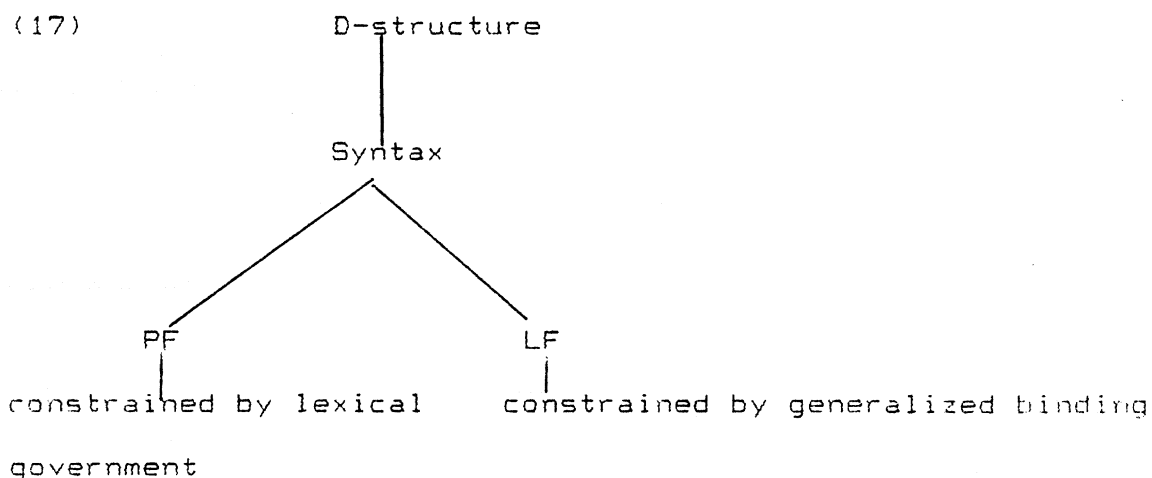
government to a modified form of antecedent government. In this section, I will try to show that lexical government should play a more central role in the theory of proper government. I will present some arguments from WAHL (1987) and Koopman and Sportiche (1986) that argue for treating lexical government as an integral part of the theory of proper government. I will also draw on evidence from WAHL to show that the lexical government must apply in addition to the theory of generalized binding, which replaces antecedent government in this framework. I also assume that lexical government is a purely structural relation that applies even where there is no theta marking between the proper governor and the empty position. A last assumption is that lexical government applies at PF, while antecedent government applies in LF. A sketch of this model is given in (17). Finally, I assume that government can apply across an adjacent maximal projection, and if the category is a proper governor, proper government will apply in this domain as well¹⁴.

The definition of government is given below.

- (16a) (a) governs (b) if (a) is an X^0 , a sister of (b) or (b's) maximal projection.¹⁵

Proper government is a subcase of government and is restricted to X^0 s that are lexical.

The resulting model of the application of the two parts of the ECP is given in (17).



As a first example of how lexical government works, let us take the case of "that -trace" violations. These are cases like (18) which are unacceptable in English.

(18a)*Who do you think that left?

vs

(18b) Who do you think left?

In the structure (19a), we can see that the category that is a sister to the empty category's maximal projection is C^0 , which is not a lexical item. The Wh element in the SPEC C' is not a sister and is not an X^0 . Therefore it is not a potential proper governor for this empty category. However

the trace in the embedded subject position must be properly governed or the structure will be ruled out by the lexical government part of the ECP.

(19) [_{CP} [_{who}_i [_C'do [_{IP} you think [_{CP} _e_i [_C _e_i] [_{IP} _e_i left]]]]].

Notice however, that the WH element in SPEC seems to be able to transfer its features to the head of the complementiser. If we assume that selection is uniformly a Head to Head relation as proposed in Chomsky(1986), then we should assume that the WH in the specifier in a case like (19) or (20) can transfer its features to the head of CP, so that the selectional restrictions of 'wonder' are met.

(20) You wonder [_{CP} _{who}_i [_C'e [_{IP} _e_i left]

Notice also that in relative clauses like (21), "that-t" effects are cancelled.

(20b) The man that left is a nice guy.

Following Pesetsky(1981), this has been interpreted to mean that the 'that', which in normal cases cannot function either as a binder or antecedent governor, can somehow take on these features in this case. Pesetsky(1981) claims that this is because there is a process of index transfer between the head of the relative clause (perhaps through the relative

operator) and the complementiser. We can interpret Pesetsky's story as claiming that if a category bears the index of another category, it can serve all the functions served by the category with which it is coindexed. In the case of relative clauses, it can serve as a binder.

Following this line of reasoning, we claim that if the Wh element, which is a lexical category could transfer, or copy its index onto the complementiser, then the complementiser would be treated as a lexical category. Since it is intrinsically a head, it would thus be a lexical head and could serve as a lexical proper governor for the subject trace in (18).

The ungrammaticality of a case like (18a) then comes down to the question of why transfer of the WH element to the head of CP is disallowed in this case. Recall that, independent of these constructions, English bans CPs that have lexically filled specifiers and heads as in (21).

(21)* I know who that John likes.

Let us express this restriction as the filter (22)

(22) * [_{CP} CP (a_i) [_C (b_j) ... where b is lexical.

Under normal conditions, only the index of the head of CP will percolate to the X max projection of this category. If the index of the specifier is transferred to the head position though, given standard percolation conventions, it will also percolate to the CP.

Thus (18a) is ruled out because, in order for the head of CP to lexically properly govern the subject position, the WH element must transfer its features to the head of CP. However, transferring these features will produce a structure violating filter (22). Transferring the index to the head in (18b) will not produce a violation because the head of the CP is non-lexical and thus the resulting structure does not violate filter (22). A structure like (23) is also possible because, since the trace of the object is independently lexically governed by the verb, the transfer of the index of the Specifier CP to the head of CP is not required and so the resulting structure will not violate filter (22).

(23) What did you think that John liked

(23b) [_{CP} what [_{did}]_i you think [_{CP} e_i [_{that}]_j] [_{IP} John liked e_j]

This theory predicts that if a language allows its complementisers to be doubly filled, it should also allow

'that - trace' violations. This prediction is confirmed in Germanic languages like Dutch as shown by (24)

- (24) Wie denk je [_{CP} e_i [dat [_{IP} e_i gekomen war
 who think you that came
 Who do you think that came

- (25) Wie denk je dat hem gezien heeft
 who think you that him saw has
 who do you think that has seen him? ¹⁴

Another case of lexical government in action comes from Kayne (1981) and Stowell (1981) who showed that the deletion of complementizers is subject to lexical government. They took Comp to be the head of a clause and showed that the governed complementizers in (26) may be deleted, but not the ungoverned complementizers in (27). Since these complementizers are not coindexed with NPs, antecedent government is not relevant for these cases.

- (26) a. it was [_{CP} apparent [_{CP} (that) Kay left]]
 b. the book (that) Kay wrote arrived
 c. it was obvious (that) Kay left

As (27) shows, lexical proper government seems to hold under a condition of adjacency. There cannot be an intervening

lexical item between the proper governor and the category that is properly governed.

- (27) a. it was apparent yesterday *(that) Kay left
 b. the book arrived yesterday *(that) Kay wrote
 c. *(that) Kay left was obvious to all of us
 d. Jay believes, but Kay doesn't, *(that) Ray is smart

The same requirement extends to wh words in Comp. Assuming that non-restrictive relatives and some factive constructions have the structure of (28a), where the relative clause or factive "complement" is outside the head NP, and that restrictives have the structure of (28b), we account for the non-deletability of wh words in non-restrictive factives by claiming that a deletion site in Comp must be lexically governed, in this case by the head noun. So the ungoverned wh word in (29a) or (29b) may not be deleted, unlike the governed who in a restrictive relative (29c). Similarly, the relative pronoun in restrictive relatives like (30) may not be deleted because it is not governed by the head noun: an extra maximal projection (the PP and NP indicated) separates the head from the wh word.

- (28) a. [NP] S'
 b. [det_{NP} [N' S']]¹⁷

- (29) a. Jay saw Kay, *(who) I admire
 b. The claim *(that) Mary loved Fred ¹⁹
 c. Jay saw a guy (who) I admire
- (30) a. the guy _S [_{PP} [to *(whom)] I wrote]
 b. the guy _S [_{NP} *([whose) house] I bought]

In order for this account to work, we must assume that even when a complementiser is unrealised at s-structure, it is present in the PF component, where lexical government applies. Let us assume that a category is visible in PF if it is lexical or if it is indexed. Let us further assume that "deletion" leaves an indexed element that is visible in PF, at least at the level where lexical government applies.¹⁷ Let us also assume when a category is moved it leaves an indexed element.²⁰

Let us now consider a case that shows that the condition of head government should apply at PF and not at S-structure. We consider the case of preposition stranding, which occurs freely at LF but is generally not possible in the syntax. Syntactic stranding in modern English must involve some marked property, as noted by many writers: perhaps a reanalysis process along the lines of Hornstein and Weinberg (1981) or a process extending government across a PP

node along the lines of Kayne (1981), so that the trace is lexically governed. Whatever that process is, it must apply after stylistic rules which can affect government relations, for example the permutation rule relating speak to Jay tomorrow and speak tomorrow to Jay. If after these stylistic rules, the V governs the PP, then the marked process permits the V to govern the NP object of the preposition, permitting (31a) but not (31b for the same reason that 31c with a non-governed PP is not permitted).

- (31) a. who_i will you speak to e_i tomorrow?
 b. *who_i will you speak tomorrow to e_i?
 c. *which concert_i did you sleep during e_i?

This shows not only that a condition of lexical government must apply after at least some stylistic rules, hence post S-structure. These examples also show that imposing a condition of lexical government at LF leads to some complications. Since prepositions may be stranded freely at LF, then their traces would have to be lexically governed at LF. A case in point is,

- (32) a. which man said he would speak to which woman
 b. which man said he would speak tomorrow to which woman
 c. which man slept during which concert²¹

Generalized Binding

Given that the lexical government requirement applies at PF, it must be supplemented with a locality condition that will also apply to LF movement. In this section, we consider the distribution of adjuncts and arguments within the Generalized Binding framework. ²²²

GEBI requires that anaphors be locally bound in their domain, if they have one. The theory is "generalized" in that this binding theory applies to the elements that are part of both A and A-bar chains. The binding theory is a generalized version of one proposed in Chomsky (1981) and given in (33). In order to be bound, a category has to be coindexed with a c-commanding element in its domain.

(33) (where $X = A$ or A')

- A. An X-anaphor must be X-bound in its domain.
- B. An X-pronoun must be X-free in its domain.
- C. R-expressions must be A-free.

We also follow Aoun (1985) in claiming that the domain for Binding consists of the first NP, S or S' that contains an accessible SUBJECT, where a SUBJECT is the most prominent nominal element (NP or AGR) within that domain.²³ We also adopt Aoun's notion of accessibility.

(34) (a) is accessible to (b) iff (b) is in the c-command domain of (a) and assigning the index of (a) to (b) would violate neither the i within i condition²⁴ nor Principle C of the binding theory.

Before considering some cases, let us finally discuss how categories move into COMP and the conditions under which binding of WH traces takes place. We assume that CP consists of a specifier which is the landing site for a WH element. Thus, in the normal case, a WH element will move into Spec CP from which it c-commands every thing in the clause that it dominates and so can bind a trace in the clause. (36a) requires that the binding condition to be satisfied, but (36b) does not.

(36a) [_{CP} who_i [_{IP} e_i left]]

(36b) [_{CP} what [_C-did [_{IP} John like e_i]]]

In both of these cases the wh-trace is an A' anaphor. As a WH element it is also subject to Principle C of the Binding

theory. These "A/C" anaphors have domains, only if they have accessible subjects given the definition of accessibility cited above. (36a) has the structure (37a), and AGR as its accessible subject.

(36a) [_{CP} who_i [_{IP} e_i AGR left]]

The object trace shown in (37a) has no domain because it has no accessible subject. The NP in subject position is not accessible because coindexation between it and the object trace leads to a violation of Principle C. AGR also cannot count as an accessible subject, because by standard conventions, it is coindexed with the subject position. Therefore coindexing the object trace with AGR will also lead to a Principle C violation.

(37a) [_{CP} what_i [_{IP} did John_j [_{AGR} [_{CP} like e_j]]]]

We also assume that a trace in COMP inherits the status of the category at the foot of its chain. If the foot of the chain is an A/C anaphor, then the trace in COMP will be an A/C anaphor. More importantly, we assume that if an A' anaphor must be bound as in (36a), it must be strongly bound at LF. That is, its chain must be linked to a c-commanding lexical operator at LF. If, the trace does not have a domain, then its chain must still be linked to an operator,

but that operator need not c-command or bind an element in the chain.²⁵

These assumptions are natural, and are crucial for our treatment of adjuncts.²⁶ Let us first reconsider (8), repeated as (38).

(38) [_{CP} Why_i did [_{IP} Nory [_{CP} [_{CP} mow the grass] e_i]]].

The first question that we must ask is whether the adjunct 'why' has a domain or not. We follow Aoun (1986) and WAILL (1987) in claiming that it does because, unlike the object trace, the adjuncts 'why' and 'how' are non referential. Following Huang (1981) and Aoun (1985) we claim that there is a syntactic notion of referentiality.

In the unmarked case "referential elements" tend to be noun phrases and we tend to find a set of deictic pronouns for these elements. Notice that wh elements like 'who', 'what' and 'which' have corresponding deictic pronouns ('he', 'it', 'this', as do the temporal and locative phrases 'where' and 'when' ('here', 'there', 'then' and 'now'). There are no corresponding deictic pronouns for 'why' and 'how'. The grammar seems to treat persons, times, things and places differently from reasons and ways of doing things. It seems possible to pronominally pick out the former but not the latter. We mark this distinction by dubbing the former

"referential". This description is intended to have grammatical, not metaphysical, significance.

Since reasons and ways of doing things are non-referential, the traces of 'why' and 'how' are not subject to Principle C of the Binding theory. Therefore, unlike traces that are A/C anaphors, the traces of 'why' and 'how' have a domain and must be bound in the clause that immediately dominates them. In (38), the adjunct trace is bound and thus the sentence is acceptable.

Contrast this case with the unacceptable (11c) and (d), repeated as (39)

- (39)*a. How do you wonder whether John fixed the sink.
 *b. Who wonders whether John fixed the sink how.

Since GEBI applies at LF, both structures must conform to it. The structure for (38a) is given in (39c).

- (39) c. [_{CP} how_i [_{do}] [_{IP} you [_{VP} wonder [_{CP} whether [_{IP} John [_{VP} [_{VP} fixed the sink] e_i]]

The adjunct is not bound in the lowest clause as another WH element is filling this position. Therefore the structure is ruled out by GEBI.²⁷

In (39b), the WH in the matrix clause moves into the SPEC CP in the syntax.²⁹ We assume that WH elements must be adjoined to the NP occupying the SPEC CP position, yielding a structure like (40), which will subsequently be interpreted as an absorption structure.

(40) [_{CP} [_{NP_i} [how_i] [_{NP_j} who_j] [_{CP} [_{IP} e_j [_{VP} wonders [_{CP} whether [_{IP} John [_{VP} [_{VP} fixed the sink] e_i]]]]]]]]³⁰

The underlined trace in (40) is an A-anaphor. However it is not strongly bound at LF because the quantifier that should ultimately bind its chain is adjoined to the maximal projection NP_j. Thus, while NP_j c-commands the rest of the structure, it prevents the adjunct from c-commanding and strongly binding the adjunct trace in the underlined position.³¹

Since the adjuncts 'when' and 'where' are referential A/C anaphors, they have no domain. Thus we expect them to be able to take a wide scope interpretation in (14), repeated as (41).

(41) Who asked what happened when

[_{CP} [_{NP_i} when_i] [_{NP_j} who_j] [_{IP} e_j asked [_{CP} what_i [_{IP} e_i happened e_j]]]]³¹

The two subject traces both have accessible subjects and must be bound. The adjunct has no domain and so it can simply be linked to the non c-commanding WH in the matrix.

Next, let us discuss the way that GEBI handles binding into noun phrases and compare it to the Barriers approach. We will consider relative clauses, noun complement constructions and movement from the subject position. For the moment we will not discuss violations as they relate to bounding theory, but will discuss those movements that also invoke a Generalized Binding or antecedent government violation.

*(42) How do you know a man who fixed the sink?

First consider how (42) is handled within the Barriers framework. On the assumption that categories cannot adjoin to NP or CP, movement out of the relative clause will produce a violation of antecedent government, the only disjunct of the ECP applicable to adjunct traces. The antecedent of the underlined trace is separated from it by two barriers, the CP and the NP which inherits barrierhood from the CP. The structure is given in (43)

- (43) [_{CP} how_i [_{do} [_{IP} you [_{VP} e_i [_{VP} know [_{NP} a man_j [_{CP} that_j [_{IP} e_j [_{VP} e_i [fixed the sink] e_j]]]]]]]]].

We follow the analysis in WAHL to handle these cases. Recall that an adjunct is a pure A-anaphor, and so must be bound in the domain of its subject. Since there is already a relative clause operator in the SPEC CP, the adjunct must adjoin to the NP in Spec CP. However, it will no longer c-command the adjunct in the embedded clause and so the structure is ruled out.

- (44) [_{CP} how_i did [_{IP} you [_{VP} know [_{NP} e_j [_{NP} a man_j [_{CP} [_{NP} e_i [_{NP} who_j]] [_{IP} e_j [fixed the car] e_i]]]]]]]]³²

In the noun complement case, things are somewhat more complicated for the Barriers analysis. Since the complement is l-marked, it does not count as a barrier for antecedent government, and if the NP is in object position, it is also l-marked by the verb. Since no barriers are crossed, movement should be possible either in syntax or in LF. To handle cases like (45) Chomsky proposes a "minimality" condition that works to block antecedent government even in cases where no barriers are crossed if there is an intervening potential governor. Thus the head NP in a noun complement structure blocks antecedent government from

outside of the noun complement. We will discuss this condition more formally in chapter 6. For now the intuition will suffice. In (45), the VP- adjoined trace inside the noun complement structure cannot be antecedent governed because the intervening nominal head is a possible governor and so protects this trace from outside government. Notice in this case as well, that Chomsky must not allow the intervening trace to delete before LF.

(45) *how do you know a plan to fix the car

(45b) [_{CP} how_i do [_{IP} you [_{VP} e_i [_{NP} a plan [_{CP} PRO [_i to fix the car] e_i]]]]

The GEBI framework rules out (45) as it does (42). The most prominent nominal element in the noun complement is the head N'. Therefore this counts as a subject and so the adjunct must be bound inside the NP. All the intermediate traces of the adjunct are A-anaphors and so they must be bound inside the NP. Therefore any subsequent movement outside of the NP will be blocked because the adjoined trace that it would leave behind would not be bound in the structure. The offending trace is underlined in (46).

(46) [_{CP} how_i do [_{IP} you know [_{NP} a [_{N'} plan [_{IP} e_i [_{IP} PRO to fix the car e_i]]]]]]³⁰

Movement from noun phrases will be discussed in some detail in chapter 6. For the moment, we note that any category that needs to be antecedent governed cannot move from subject noun phrase position in the syntax given the Barriers framework. The NP in subject position is not L-marked and cannot be adjoined to for reasons given above. Thus this NP will always block external antecedent government.☞

We will explore the divergent empirical predictions that the Barriers and Wahl theories make about movement from a subject NP in chapter 6.

Having discussed GEBI and lexical government separately, I will present a case that argues that structures must satisfy both lexical government and generalized binding.

This case is taken from Koopman and Sportiche(1986).

Koopman and Sportiche notice that adjuncts can only be moved in the syntax in Vata, a West African language, if a special morphological suffix is appended to the verb.³⁵ This applies both to reason and manner adverbs.

(47a) Yeso n dido suo la
 how come you cut-suf tree-det WH
 why did you cut the tree

*(b) Yeso n di suo la
 why did you cut the tree

(48a) so n na ka suo dido
 like-this I said I-fut tree cut
 It is like this that you said you cut a tree

*(b) so n na ka suo di
 like-this I said I-fut tree cut

Assuming that this morphology acts as a lexical proper governor and assuming that the ECP was satisfied if either disjunct of this condition was met, we would predict that long movement of an adjunct would be possible from a position governed by verbs with this special suffix. This prediction is not borne out as shown by (49).

*(49) Yeso n ylanyni ze n didodido- bo la
 how you wonder thing you cut+morph Rel

Long movement of non-adjuncts is possible, as shown by
 (50).

(50) alo n nI ze n ka bo nyE yi la
 who you NEG-A thing you fut Rel give know WH

Who don't you know what you will give to?

Long movement of PP adjuncts is also impossible as shown by
 (51).

*(51) yi gbu n nI ze n ka bo nu yi la
 what cause you neg-A REL you FUT REL do know WH

Why don't you know what to do?

We can explain this pattern by assuming that NP adjuncts are subject to both lexical government and to generalized binding. Both categories are subject to generalized binding. The adjuncts, being non-referential

elements, fall under Principle A of the Binding theory. PPs, also being non referential, fall under Principle A and so both types of cases have to be bound in the domain of their first accessible subject, barring long movement.³⁷

A further case that argues for applying both lexical government and generalized binding comes from reconsideration of super raising structures like (52).

(52) John_i seems it appears e_i to be intelligent.

Since we have divorced theta marking from our notion of lexical government, the trace in the lowest clause will be lexically governed by "seems". The sentence is still ruled out by generalized binding because the trace, which is an A-anaphor is not bound in the domain of its first accessible subject; the medial clause. Several authors have questioned this analysis because sentences like (52) sound much worse than sentences where binding of an overt anaphor is blocked in similar constructions. We attribute this extra degree of ungrammaticality to an additional violation of the theta criterion along the lines of Haik (1985). Haik suggests that theta roles can only be assigned to well formed chains. Assuming that theta roles are assigned to chains (see Chomsky (1981)) it is natural to assume that if a chain cannot be formed given independent principles of the grammar, then this

object will not be available for transmission of a theta role to its head.³⁹ Notice that this idea also accounts for why the structures in (53) sound much worse than the corresponding structures in (54).

- (53) *a. The men_i were believed e_i were happy
 *b. Fay was believed pictures of e were on sale

- (54) *a. The men believed each other were happy.
 *b. Who did Fay believe pictures of were on sale.⁴⁰

This assumption or something like it seems to be needed in the Barriers framework as well, on the assumption that lexical anaphors are actually moved in LF, leaving empty categories subject to the ECP. (55) will then have the same structure as (53a), but nonetheless, (55) is less deviant.

- (55) The men believed each other were handsome.
 b. The men_i each other_i believed e_i were handsome.⁴¹

Consideration of a case like (56) suggests that dispensing with the i inside i condition is quite problematic even within the Barriers framework.

- (56a) The men_i believe [_{CP} that [_{IP} [_{NPE} some reviews of [_{NP}_j each others_i books]] are in the library]].

In this case movement of 'each other' at LF will leave a non properly governed trace. Lexical government is not applicable because 'review' does not theta mark the determiner position of NP_j. 'Each other' is contained in NP_i, which is a barrier . Moreover, the anaphor cannot adjoin to this NP as this is completely disallowed in the Barriers framework. We cannot dispense with the i inside i condition unless we treat "binding" possibilities of subjects of tensed clauses as ECP phenomena because, unless we assume that the AGR sets up an opaque domain, tensed clause cases will not be ruled out under Chomsky (1986)'s version of the binding theory. If we treat these cases by moving the lexical anaphor in LF, we seem to rule out acceptable cases but if we treat these cases as binding theory phenomena, then we will have to reintroduce AGR as an accessible subject or we will expect to be able to bind an anaphor in a tensed clause from outside of that clause, which is impossible.

4.3 Conclusion:

In this chapter, we have considered two approaches to the ECP. We have tried to suggest that the Barriers framework captures the facts, but only at the cost of making some rather unnatural assumptions regarding the levels at which categories can be governed and the types of categories that elements can be adjoined to. We have presented an

alternative which takes the referential status of adjuncts into account and so easily distinguishes between adjuncts like 'why' and 'how' vs. 'where' and 'when' in terms of their syntactic distribution. We have also presented a case for treating lexical government as a PF condition which applies in addition to GEBI or antecedent government. In the next chapter we will discuss what constraints on overt syntactic movement would look like in these frameworks and critically evaluate the two approaches. In chapter six, we will provide further tests of these approaches as alternative versions of the ECP.

Footnotes

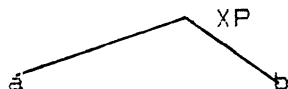
1. Henceforth, this work will be referred to as WAHL.

2. Actually, this statement is a bit misleading because Chomsky(1986) presents two alternatives, one as presented in the text, and one where lexical government is eliminated and the ECP is reduced to antecedent government. We will discuss this latter alternative below, suggesting that this runs into a variety of empirical problems. For the moment however, we will concentrate on the disjunctive alternative for proper government.

3. Our discussion of gapping in the previous chapter crucially relies on treating S or IP as a maximal projection and we will continue to make this assumption for the rest of this thesis.

4. This is a somewhat simplified account of the definition given in Chomsky(1986) pg.25

5. The notion of "m-command" comes from Aoun and Sportiche(1983) and obtains when every maximal projection dominating (a) dominates (b) in the structure



6. As mentioned above Chomsky(1986) tries to reduce proper government to antecedent government in the latter part of this work. We will discuss some problems for this directly and thus will concentrate on the version of the Barriers framework that retains both disjuncts of the ECP.

7. Chomsky(1986) correctly observes that super-raising from NP should not be blocked as a simple subjacency violation because the results are much worse than long WH movement in similar structures like (a)

(a) What do you wonder who saw.

8. Both definitions are from Chomsky(1986) pg. 9.

9. This assumption is justified in May(1985). It is crucial in allowing subsequent movement from the adjoined VP to COMP, because if the VP contained the adjoined trace, it would count as a barrier and block antecedent government, under the assumptions discussed above.

10. We follow Huang(1981) Aoun(1985), Chomsky (19xx) and Aoun, Hornstein and Sportiche(1981) in claiming that Wh in situ constructions can be interpreted as double questions in English only by actually moving the WH in situ to the position containing a WH Specifier or head of CP.

11. It should be obvious that long movement from the subject position in the syntax or in LF is also barred for exactly the same reason. A relevant case is given in (a) and its structure is given in (b).

*(a) Which man do you wonder who liked

(b) [_{CP} Which man_i do [_{IP} you [_{VP} e_i [_{VP} wonder [_{CP} who_j e_j liked e_j]]]]]

12. See Zubizarreta (1982), who argues that these phrases are "adjunct theta" roles that meet the adjunct theta criterion only at LF. See Hornstein (submitted), who argues that these elements are pure modifiers whose presence is checked in PF. Under either assumption, they should behave differently from true arguments that must be present at s-structure in order to satisfy the Projection Principle.

13. See also Hornstein and Lightfoot (forthcoming) for further empirical arguments against this proposal from the grammar of Polish.

14. In this we follow Kayne(1981), who allows case marking to apply across one maximal projection in French.

15. Given this definition, a verb will govern an NP in (a). Notice that the preposition also governs this position. We will assume, loosely following Chomsky, that government should be a unique relation and that the closer potential governor blocks government by the farther governor. We will use Chomsky's definition to obtain this result:

(a) [_{VP} V [_{PP} P [_{NP} e]]]

(b) In the configuration

...(a)....[x...(x)....(b)]

(a) does not govern (b) if (X) is a projection of (x) excluding (a).

Chomsky(1986) pg. 42

16. This correlation was first pointed out by Koopman(1984) and others.

Sobin(1987) points out that there are dialects of

English that allow "that-t" violations in non-relatives. Sobin assumes ala Pesetsky, that in these cases, the index of the WH-trace is not simply transferred to the complementiser, but is "fused" with it, creating a category with a unique index and a unique set of features. In marked dialects which allowed this extension of "fusion" to non-relative clause structures, we would predict that there might be no correlation between violations of the "doubly filled COMP" and "that-trace" filter. Sobin does not discuss whether English speakers who allow that-trace violations also allow doubly filled COMPs, so his data cannot decide the issue. Interestingly, neither Dutch nor English allows this kind of "fusion" when the complementiser 'whether' replaces 'that'. This would be predicted if we assumed that the WH complementiser moved into SPEC CP at S-structure, as argued in Chomsky(1973). In this case, the WH complementiser would not be in the appropriate position to properly govern the subject trace, whether or not fusion had occurred.

17. Williams(1986) provides independent evidence that this is the structure for restrictive relatives. He notices that (a) but not (b) is acceptable.

- (a) pictures of each other_i that the men_i like are on sale.
 *(b) each other's_i pictures that the men_i like are on sale.

He attributes this difference in acceptability to the fact that since the relative clause is predicated of the N', this N' can be reconstructed within the relative clause. Assuming that the relative is not predicated of the determiner, we would not expect reconstruction to apply in (b), thus predicting its unacceptability.

18. Crucially, we follow Kayne(1981) and Stowell(1981) in claiming that the attribution of "noun-complement construction" to cases like (29b) is a misnomer. These authors claim that there is a difference in the semantic relation of a tensed and untensed clause to the head noun in these constructions, with a tensed clause in an appositive relation to the head, and an infinitive in a complement relation. Given this semantic difference, we would expect the two constructions to have different structures. Evidence for this claim, taken from Stowell(1981) shows that the tensed complement in effect identifies the same entity as the head of a "factive" NP, while the infinitive functions as a true complement. As such we expect that only tensed clauses can appear in "identity statements" with the head of the NP while infinitives cannot.

- (a) Andrea's claim that she was insecure

(b) Andrea's claim to be insecure

(c) Andrea's claim was that she was insecure

*(d) Andrea's claim was to be insecure

See Stowell(1981) chap 6. for further arguments along these lines.

Chomsky(personal communication) claims that this argument is inconclusive because there are cases where the tensed "complement" in a noun complement construction does not seem to establish an identity relation. Such cases are given in (e).

(e) The proof that four is divisible by two.

Notice that this same "nonidentity" reading occurs in clear appositive constructions.

(f) Four, which is divisible by two...

Stowell and Kayne's claim then is really the claim that tensed constructions are parallel to appositives, in that they identify but do not necessarily uniquely identify the head with which they are construed.

19. This assumption deserves further comment.

Chomsky(personal communication) points out that our notion of PF cannot be equated with a notion where PF is taken to be the level at which only those categories that are pronounced are represented because our empty, indexed categories are certainly not pronounced. Notice however, that since a variety of PF processes like contraction, refer to empty categories (see Lightfoot(1977, 1986), Jaeggli(1980) and Aoun and Lightfoot(1984) for details), we have independent evidence for a level of representation where both empty and lexical categories are present and which can feed PF processes. Thus this account only adds the assumption that both indexed and case marked empty categories can be visible at this more abstract level.

20. We must crucially assume that PRO need not be indexed at s-structure or at PF because it can appear in ungoverned positions. We will see the effects of this lack of indexation in chapter 7.

21. Chomsky (personal communication) claims that we could disallow movement from the extraposed structures cited above simply by assuming that the verb preposition sequence is reanalysed at s-structure. This would block extraposition

because there would be no PP to extrapose. Sentences like (a) where the NP 'advantage' can be passivized even though it is internal to a reanalysed structure show that reanalysis does not block movement from reanalysed constituents and thus we would not expect extraposition to be blocked even if these structures are reanalysed.

(a) who was advantage taken of

We will see below that there are theory internal reasons for claiming that lexical government applies in PF. If we claim that the ECP governs the deletion of a complementiser then it must apply at the level when these complementisers are deleted. However, we must claim that adjuncts in complementiser position and WH elements linked to the subject position are present at LF or else we will induce a violation of generalized binding at LF as we will see below. Examples are given in (b) and (c).

(b) The reason (why) John arrived

(c) The man(who) you thought would win the race.

22. Henceforth, we will refer to Generalized Binding as GEBI. The theory to be discussed here is a slightly modified version of the theory presented in Aoun(1981)(1985) and WAHL(1987)

23. In order to be "bound in its domain" an antecedent must bind its argument in the first domain in which the argument could be bound. For a WH operator, the first potential antecedent position is in the c-commanding Spec CP position. Thus, the first domain is the CP. For a non WH element, the first potential antecedent position is either adjoined to IP or within IP. Thus, the first domain is the IP.

We assume the definition of c-command provided in Aoun and Sportiche(1983) except that, following Reinhart(1983), Aoun and Hornstein(1985) and Weinberg and Hornstein (1987) we assume that a phrase may c-command material that it contains.

(i) (a) c-commands (b) iff every maximal projection that dominates (a) dominates (b).

24. The i within i condition says that:

*[_(b)... (a) ...] where (b) and (a) bear the same index.

25. It should be clear that if the foot of a chain has no domain, then all the traces intermediate between it and its head have no domain. This is because if, any link in a chain were to be coindexed with any c-commanding subject, the foot of the chain would also be so coindexed, inducing a Principle C violation.

26. Chomsky (personal communication) suggests that constraining the distribution of anaphors in this way is inelegant, particularly in comparison with the approach to the distribution of these categories outlined in Chomsky (1986). The approach to the binding theory outlined in this work governs the distribution of anaphors in tensed clauses by the ECP, rather than by the binding theory. In such a theory AGREEMENT does not count as a potential binder, and the *i* within *i* condition is dispensed with. It seems however that under the Barriers analysis of the ECP, this idea runs into trouble. If we try to reduce all cases of the ECP to antecedent government, then a case like (a) which moves the quantified expression from NP internal position, should be an ECP violation, on a par with (b).

(a) The men thought pictures of each other were on sale.
 (a') The men_i each(other)_i thought [_{CP} [_{IP} [_{NP} pictures of e_i(other)]] were on sale]]

*(b) The men thought each other were nice.
 (b') The men_i each(other)_i thought [_{CP} [_{IP} [_{NP} e_i] were nice]]

In both cases, the trace is separated from its antecedent by 2 barriers, NP and IP. We will have more to say about these cases in Chapter 6.

27. We make the standard assumption that 'whether' must occupy both the specifier and head position of CP at LF in order for the clause containing 'whether' to be correctly interpreted as an indirect question.

28. This entails non-adoption of Chomsky(1986)'s condition on vacuous movement; the claim that WH elements do not move vacuously in the syntax. Elements in the subject position do not move into COMP until LF and selectional restrictions are satisfied at this level. Rather, we claim that selection restrictions must be met at both s-structure and LF for languages that exploit syntactic movement, and at LF for languages like Chinese where WH elements move in LF. Given the VMH, we would expect that the WH in situ would be able to take narrow scope in a sentence like (a) because the embedded verb would select its complement at LF. Assuming that selectional restrictions apply at s-structure in French correctly rules out this the unacceptable reading.

(a) Jean sait que Pierre a vu qui.
 Who does John think that Pierre saw
 * John knows who Pierre saw

Kenstowicz (1986) also provides evidence against this condition using evidence from Tangale. Kenstowicz shows that sandhi phenomena apply to the last segment in a word if that

word is followed by another word in the same phrase. Sandhi occurs to a verb when followed by a postverbal object. Interestingly, sandhi does not occur when this element is a WH element. Kenstowicz explains this result by claiming that the WH object moves vacuously into the right peripheral SPEC CP position, which is the normal position for WH elements in this language. Kenstowicz also shows that only one WH element can move to this position in the syntax. This predicts that in a double question, since one WH word would have to remain in situ, it should trigger the phonological change on the verb inside its VP. This is what is observed. Given the VMH, we should predict sandhi in both cases because there would always be a WH in situ at the level where sandhi applies.

29. Recall that we follow Aoun (1986), Chomsky (1981) and Aoun, Hornstein and Sportiche (1985) in claiming that WH in situ constructions can be interpreted as double questions in English only by actually moving the WH in situ to the position containing a WH Specifier or head of CP.

30. The quantifier that was originally moved to the specifier position can bind the subject trace in the matrix clause because its lexical features can percolate its maximal projection.

The same remarks will force the adjunct in (a) to have narrow scope in a language like Chinese, as described by Huang (1982) and Lasnik and Saito (1984). Consider the contrast in (a)

(a) Ni xiang-zhidao shei shuo Lisi mai-le sheme
you wonder who said Lisi bought what

(b) Ni xiang-zhidao shei shuo Lisi weisheme mai-le shu
you wonder who said Lisi why bought the book

(a) is ambiguous, with either (WH) element taking wide scope but the adjunct in (b) can only have scope over the embedded clause. The structure for the non-occurring reading is given in (c).

(c) [_{CP} weisheme [_{CP} ni xiang-zhidao [_{CP} shei [_e shou
[Lisi mai-le shu]]]].

In this structure the adjunct is not bound in the domain of its first accessible subject and so the sentence is out.

An alternative derivation which moves 'weisheme' successive cyclically, to save the sentence from being a GEBI violation will not help matters. We claimed that selectional restrictions are met for verbs that select WH complements by having the category in the specifier transfer its WH feature to the head. If weisheme moves to the specifier position first, then subsequent movement will leave a trace in this position. We follow Lasnik and Saito in considering traces as [-WH]. Therefore, the selectional restrictions on this verb will not be met given this derivation.

As Huang (1982) points out, the subject in Chinese behaves like an object in English and can long move. We follow Aoun (1986) and WAHL (1987) in accounting for this fact by claiming that Chinese has no AGR, and therefore the subject position has no domain or accessible subject. As Aoun notes, this idea also predicts that we should be able to find reflexives and reciprocals in this position. This is in fact the case.

31. Things become more complicated in the case of overt movement. In general, it seems that there is a preference to associate adjuncts and other Wh phrases with the nearest structure clause. Thus in (a), the most natural reading associates the adverb with the matrix verb.

(a) When did you say that John left.

Also, cases like (b) are marginal. In both cases, it is possible to interpret the WH element in the upper clause, respecting the WH island condition. This seems to be the preferred option.

??(b) When did you know what to do e_1 .
 ??(b') Who did you ask whether we met e_1 .

Notice though, that when there is overt evidence of the position from which an adverbial phrase is moved, the sentences are improved, comparing favorably to simple WH island violations.

(c) Where did you wonder whether to move to?
 (d) Where did you wonder which boy came from?

Since extralinguistic factors seem to be playing a role for overt movement, LF movement seems to provide a better test of the domains in which an element can be bound.

32. We should say a few words about languages like Spanish and Hebrew where relative clause structures can contain resumptive pronouns. Let us assume that these structures do not involve movement of an operator binding a pronoun into Spec CP. Movement of an adjunct is nonetheless unacceptable from within these structures at LF. Consider (a), a translation of these cases.

*(a) [_{CP} How_i [_{NP} e_i [_{NP} a man [_{CP} e_i [_{IP} you helped him e_i]]]] repay you].

There are two possibilities here. If these languages only allow LF movement into a [+WH] Spec CP, then block movement will be blocked because the COMP is not marked [+WH] in this case. Even if this restriction does not apply, the underlined trace will not be bound in the domain of its first accessible subject, which is the head of the relative clause. Movement outside of the NP containing the relative clause head thus violates GEBI even if it is successive cyclic and at LF.

33. For further discussion of these sorts of noun complement constructions c.f. chapter 5. That A-anaphors adjoined to S must be bound within that S c.f. Aoun and Hornstein (1985).

Johnson (1987) presents evidence against treating the heads of noun complements as SUBJECTS.

He notices the contrast between (a) and (b)

(a) They read proofs that pictures of each other had been forged.

*(b) The read theorems that books about each other explained.

He claims that this contrast can be related to the fact that the underlined phrase is a complement to the head NP in (a) but not in (b). He crucially relies on the assumption that noun complement structures, with finite complements have a Det N' structure and on the assumption that an N' cannot act to set up an independent binding domain.

We have presented evidence against assuming a det N structure for noun complements above.

Moreover there is evidence that we should not relate the extension of the possible binding domain in a case like (a) to whether the post nominal clause is governed by NP or N' as Johnson's account would suggest.

Notice that picture NPs allow an extension of the binding domain beyond the embedded clause:

(a) the men thought that pictures of each other were on sale

It seems that if an anaphor is subject to the i-within-i condition it can extend its binding domain beyond the next SUBJECT:

(b) The men think that Harry said that pictures of each other were on sale

Sentence (b) is not particularly unacceptable, especially if compared with (c):

*(c) The men think that John likes each other

However, if this is correct, then it is unclear what the acceptability of (b) or (d) shows as regards the question whether the head of a noun complement structure is a subject:

(d) The men denied the fact that pictures of each other were on sale

If one accepts that it is not a SUBJECT in (d) then one would also have to claim that 'Harry' is not a SUBJECT in (b).

In addition, it is not clear that the sentences that Johnson cites are truly minimally different. The relevant noun complement sentence would be one that took an embedded question so that the effects of the WH in COMP that appears in a relative clause could be controlled for. When we contrast the appropriate pair of sentences, it is not clear that Johnson's argument goes through. Contrast (e), (f) and (g):

(e) John answered the question who ate the cake

* (f) John_i answered the question who_j those pictures of himself_j impressed e_i

?? (g) John_i answered the question which_j those pictures of himself_j suggested e_i

In our ear, the relative clause (g) sounds a little better than the noun complement (f). It certainly sounds no worse.

34. We will see in chapter 6 that this must be qualified somewhat when we consider external antecedent government of the determiner position of an NP.

35. Koopman and Sportiche(1987) note that this morphology is not allowed if the moved category is an object. They claim that the requirement of special morphology in the adjunct case, is motivated because it extends the domain of lexical proper government to configurations (like an adjunct in an adjoined VP) where the verb does not c-command the potential proper governee. This morphology is also not possible if the verb is followed by a lexical argument or adjunct.

36. Even those these sentences superficially have the form of relative clauses, Koopman(1984) argues that they act syntactically like simple indirect questions.

37. Koopman and Sportiche(1987) explain the impossibility of long movement by introducing an extra condition called the condition on long extraction. They claim that we must add this condition because an account of these data in terms of generalised binding is ad hoc. Thus they say that

"In order to derive the fact that only short movement of PP adjuncts is possible, it must also be assumed that they fall under the GEBI requirement...they must, in order to fall under GEBI, be considered non-referential. WAHL otherwise assumes that PP adjuncts ...are referential. The necessary assumption for Vata appears ad hoc."

While it is true that we must assume that PPs can differ as to whether they are treated as referential elements or not, this ambiguity is quite natural. Nouns are the referential categories par excellence. It would be natural for language learners to syntactify the concept of referentiality, equating it with projections of noun phrases that have an inherent reference. In such languages, we would expect PPs to count as non-referential categories in virtue of their syntactic form. Other language learners might take other semantic factors such as whether a category substituted for a moment of time, or location as in the case of 'where' or 'when'. Moreover, it seems more reasonable to try to assimilate this case to independently needed principles, than to create a special condition that simply governs long extraction.

Finally, it seems to be a simple empirical fact that languages differ as to whether long movement of PPs is allowed. Rizzi(1982) and Sportiche(1981) give many examples of long movement of PPs in Italian and French.

38. This account requires adoption of Chomsky's idea that the expletive is replaced by the NP it binds in cases like (a) and (b) in LF.

*(a) there were believed were two men in the room.

*(b) there seems it appears to be a man in the room.

We assume that the LF chain has 'two men'/'a man' in the subject position of the matrix sentence, a position in which it will not receive a theta role if it is not part of a well formed chain.

39. This contrast between (53b) and (54b) is due to Edwin Williams (personal communication) and is cited in WAHL(1987).

Wahl(1987) and McCloskey and Chung (1987) provide further evidence that both lexical government and binding are

needed in UG.

40. Binding Theory would not rule out (52) if the *i* inside *i* condition applied to empty categories as well as to overt anaphors for reasons discussed in Chomsky (1981). The original motivation for this condition was to block circular reference in a case like (a), where an anaphor is referentially dependent on a noun phrase that contains it.

(a) [The picture of itself]_i.

Therefore, we might expect this condition to apply only to cases where referential dependency was at issue. Since NP traces are not referential, i.e. not variables, we would not expect this condition to apply to these cases. Similar remarks apply to non bound variable pronouns. Traces of Wh movement, while variables, cannot be subject to the *i* inside *i* condition because, if such variables were coindexed with a containing NP, the structure would violate Principle C, assuming that a container can c-command categories that it contains as argued by Reinhart, Aoun and Hornstein, and Hornstein and Weinberg. Sportiche(1983) presents empirical evidence to support the conclusion that the *i* inside *i* condition only governs the distribution of anaphors.

LOCALITY PRINCIPLES IN SYNTAX AND IN PARSING

by

Amy Sara Weinberg

B.A. (joint first class honours) McGill University

(1976)

Submitted in the Department of
Linguistics and Philosophy
in Partial Fulfillment of the
Requirements of the
Degree of

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Signature of Author

Department of Linguistics and Philosophy
February 8, 1988

Certified by

Thesis Supervisor

Accepted by

Department Chairman

© Amy Weinberg, 1988

HUM.
MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

JUN 8 1988

HUM.
LIBRARIES

V6L 2

Two Notions of Bounding

In this chapter I will compare the theory of bounding in *Barriers* with one that is compatible with the WAHL framework. Both theories try to explain the domains from which extraction is blocked in terms of the theory of government, related to the ECP.

The main thrust of the bounding theory in *Barriers* is that the notion of L-marking that defines barrierhood for lexical and antecedent government also defines the class of bounding nodes for syntactic movement.

I will suggest that movement is regulated by the condition of lexical government, applying in PF. Bounding theory will force syntactic movement to proceed as a series of local steps. If the remnant of any of these steps is not lexically governed, the derivation will be disallowed. In certain cases, syntactic movement will also be constrained by GEBI, applying in IF. Even though the conditions of GEBI and lexical government do not directly interact, there will be cases where, in order for a syntactic structure to satisfy GEBI in LF, a category will have to move to a position in the syntax that will ultimately yield a violation of lexical government in PF.

5.1 Bounding in the Barriers Framework

Like earlier theories of subjacency, bounding is only permitted across one designated bounding node, but now the class of bounding nodes are taken to be co-extensive with the class of barriers. Subjacency is defined in (1).

(1) (b) is n -subjacent to (a) if there are fewer than $n+1$ barriers for (b) that exclude (a).¹

Chomsky (1986) provides both empirical and conceptual justification for this idea. One major empirical advantage of this approach is that it allows Chomsky to derive the Condition on Extraction Domains² of Huang (1982) from subjacency. In addition, Chomsky claims that this theory allows him to dispense with a substantive theory of bounding nodes. In previous theories, NP and S were stipulated as bounding nodes. The Barriers theory is supposed to pick out a set of nodes (barriers) that independently define domains of government and proper government as relevant for bounding theory.

We will discuss how this follows from Chomsky's approach, show how the standard island facts are derived by this theory, and then discuss how parametric variation in the class of bounding domains is handled by this approach.

The CED

The CED encompasses the adjunct and subject island conditions of Huang (1982) and bars syntactic movement from either a subject or an adjunct. Cases are given in (2).

*(2a) What lawnmowers are advertisements about appearing in the newspaper ?

*(b) What lawnmowers did Nory use handtools before buying?

Both subjects and adjuncts are categories that are non L-marked; an adjunct by definition because it is not a theta marked category, and the subject because, except in small clauses, the only category that governs the position is the CP, whose head is not lexical. Therefore, both categories are barriers. Since both categories are also immediately dominated by IP, movement from either structure crosses two barriers and violates subadjacency. Structures with both barriers underlined are given in (3).

(3a) [_{CP} what lawnmowers_x are [_{IP} [_{NP} advertisements [_{PP} about [_{NP} e₁]]] appearing in the newspaper

(3b) [_{CP} what lawnmowers_i did [_{IP} Nory [_{VP} [_{VP} use handtools]
 [_{PP} before [_{CP} e_i [he [_{VP} e_i [_{VP} bought e_i]]]]]]].

In order to derive the subject island part of the CED, it is crucial to disallow adjunction to NP, because otherwise the IP would not inherit barrierhood from it for the same reason that the IP does not inherit barrierhood from the VP in standard cases of movement from object position. As mentioned in the previous chapter, Chomsky stipulates that adjunction is barred to an argument position. Chomsky tries to motivate this stipulation by claiming that adjunction to an argument (CP, or NP) induces a violation of theta theory because the internal part of the adjoined structure is no longer a full category and thus might be invisible for the purposes of theta assignment.

There are several problems with this approach. Let us assume that theta marking is a relationship between a head and a maximal projection that serves as a complement as is suggested in Chomsky (1981). Normally, features assigned to a maximal projection under government can percolate down through a series of identical maximal projections to the head of the X-max. This is shown by a case like (4) where case is assigned to the NP dominating the head of the appositive construction and then can percolate to the head.

(4) I [_{VP} saw [_{NP} [_{NP} John, [_{CP} who Bill likes]]].

Thus we must claim that for some unknown reason, theta assignment does not obey standard percolation conventions.

Secondly, the approach seems too weak to block all of the relevant cases of movement from "non-argument" positions. We can show that we must block movement to NP even in cases where theta assignment does not seem to be involved. We can move a heavy noun phrase in (5a) to a non L-marked position as shown in 5h.

(5a) I composed a long letter for Mary to write to Sue yesterday.

(5b) I composed yesterday a long letter for Mary to write to Sue.

Notice however that further movement from within the heavy noun phrase is disallowed as shown in (5c) which has the structure (5d).

*(5c) Who did you compose yesterday a long letter for Mary to write to

(5d) [_{CP} Who_i did [_{TP} you [_{VP} e_i [_{VP} compose e_j yesterday] [_{NP} e_i [_{NP} a long letter [_{CP} for [_{TP} Mary to [_{VP} e_i [_{VP} write to e_j]

We assume with Chomsky that 'who_i' cannot adjoin to the CP inside the moved heavy NP because this would trigger a theta criterion violation. However, the heavy NP itself is not a position for theta role assignment. Rather the theta role is assigned to the trace in the A-position from which the heavy NP

moved. The heavy NP functions as a quasi operator³ and so we would expect to be able to adjoin to this category. Thus, movement out of the adjoined NP crosses only one barrier. Subsequent movement proceeds as would standard movement from object position. Therefore, we incorrectly predict that (5c) should be acceptable.⁴

Similar remarks apply to so-called "bare NP adverbs". Movement from these categories is also disallowed in a case like (6b).

(6a) John got drunk the day that Bill married Sue.

*(6b) Who did John get drunk the day that Bill married.

Again, adjunction to the internal CP is disallowed because the CP is an argument. Adjunction to the bare adverbial should be allowed though because this category is at most getting an "adjunct theta role", which Zubizarretta (1982) argues is only checked at LF. By LF however, the adjoined trace can be deleted and the adjunct theta role can be assigned to the structure. The structure for this sentence is (6c)

(6c) [_{CP} who_i did [_{CP} John [_{CP} e_i [_{VP} get drunk] [_{NP} e_i [_{CP} the day [_{CP} that [_{CP} Bill married e_i]]]]]]]]

In order to rule out these cases within the Barriers framework, we must claim that adjunction to NP is simply disallowed. However this means that The Barriers framework cannot simply use the barriers defined for government and proper

government as bounding nodes. Rather, bounding theory picks out NP and CP as particularly relevant. This means that we have simply shifted substantive restrictions from the class of bounding nodes to the class of adjunction sites.

Another problem is that careful inspection of (3b) shows that we must also bar adjunction to PP because otherwise the IP would not inherit barrierhood. Given that the PP is not an argument however, we cannot use the justification provided by Chomsky to block adjunction in this case. We will see below, that while movement from some adjuncts is acceptable, movement from a tensed adjunct can yield quite deviant results as shown by (7).
 *(7) Who did Jack leave London before Harry spoke to.

The tensed/nontensed distinction cannot be incorporated into the treatment of this parameter if we can adjoin to PP in these cases. Again it seems that we would prefer to disallow movement to this position, but again we cannot justify this ban in terms of theta theory.

Barring adjunction to arguments is also a crucial component of Chomsky's derivation of the island constraints. Let us consider how movement is blocked from relative clause structures like (8).

*(8) [_{CP} Which woman_i do you like [_{NP} the girl [_{CP} who_j
 [_{IP} e_j [_{VP} e_i [_{VP} saw e_i]]]]]]]

The relative clause in (4) is not L-marked by the NP. Therefore it is a barrier and since the underlined NP immediately dominates the CP, barrierhood is transferred by standard conventions. Movement thus crosses two bounding nodes. This, of course would not be the case if we could adjoin the trace inside the relative clause to the CP, for the reasons discussed above. Observe that the CP is not an argument of the head of the relative clause. However, we will show below that an empty operator inside a relative clause must be predicated of the head of the relative and since predication is under government, adjunction is barred for independent reasons in this case.

We will continue our discussion of bounding theory within the Barriers framework below. We will discuss Chomsky's treatment of noun complements in Chapter 6. Parametric variation in the Barriers framework is discussed in section 5.2. The distribution of parasitic gaps is discussed in Chapter 7.

5.2 Bounding in the WAHL Framework

As we've seen above, Chomsky assumes that every maximal projection (except IP) counts as a potential barrier for movement. There are two reasons to make this assumption. First, Chomsky

tries to unify bounding with antecedent government. We instead have claimed that antecedent government is a subspecies of binding. Therefore, there is no a priori reason to expect all maximal projections to be bounding nodes.

Secondly, Chomsky claims that we can eliminate a substantive and particular theory of bounding nodes from UG. We have seen above that the substantive theory of bounding nodes is simply replaced by a substantive theory of adjunction sites. Therefore, we lose nothing by claiming that IP, CP and NP are the relevant nodes for bounding. At the end of this chapter we will provide some reasons for why just these nodes are relevant for subjacency phenomena.

In chapter 3 we presented functional motivation for treating subjacency as a condition on S-structure. Given that the results of S-structure movement feed the PF component and that there is no deletion at this level, all empty categories in a chain formed by movement will have to conform to lexical government at PF. The results of syntactic movement must also feed IF and meet the conditions of GEBI. We will try to show that syntactic island effects fall out from the fact that syntactic derivations must meet the locality conditions imposed by subjacency at S-structure, lexical government at PF and GEBI at IF.

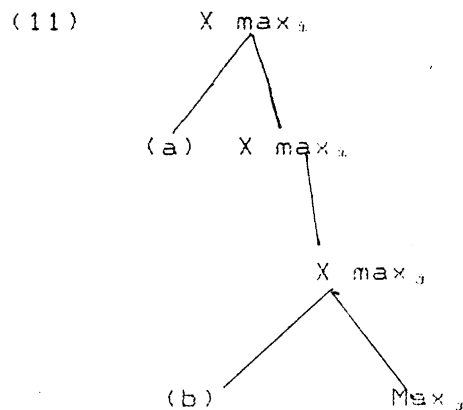
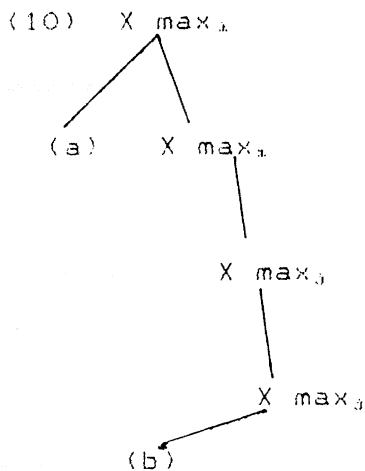
First let us define subjacency:

(9) (a) is subjacent to (b) iff (a) is separated from

(b) by at most one bounding node that contains (b) but does not contain (a).

Our definition of containment comes from May (1985).

May argues that a category does not contain an element unless all parts of a category contain that element. Thus in (10) the two nodes (a) and (b) are separated by two bounding nodes, while in (11), the x_{\max_j} does not contain (b) and so (a) and (b) are separated by only one bounding node.



(where $X\text{-max}_i$ and $X\text{-max}_j$ are bounding nodes)

We further assume that adjunction is free. In particular, a category can adjoin to any maximal projection. We will see, however, that in many cases such adjunction will lead to violations of lexical government.

Last of all, we assume that bridge verbs have the special property of voiding a category's status as a bounding node but only for elements that are governed by these verbs.²⁷

Let us begin with the simple case of COMP to COMP movement in a language like English where CP, IP and NP are bounding nodes.

(12) [_{CP} Who_i do [_{IP} you [_{CP} think [_{CP} e_i [_{IP} Bill liked e_i]]]]]

In this case the verb can move directly to the Spec CP position of the embedded clause, crossing only IP. The bridge verb 'think' allows the CP not to count for the trace in Spec CP that it governs. Thus the next movement is also allowed.²⁸

We must also deal with the so called S/S' parameter. Rizzi (1982) and Sportiche (1981) proposed that Romance languages took NP and CP as the relevant nodes for bounding theory, citing the acceptability of examples like (13) and (14).

(13) [Combien_i as [_{IP} tu mange [_{CP} e_i de gateaux]]

How many cakes did you eat

(14) Tuo fratello, a cui mi domando che storie abbiano
raccontato era molto preoccupato

Your brother, to whom I wonder which stories they told,
was very troubled."

Recently, the idea that this is a truly language specific parameter has come into some question. Both Rizzi and Sportiche claim that movement across a CP is much better when the CP is not tensed. Many English speakers also seem to allow movement across a CP in these cases as shown by (15).

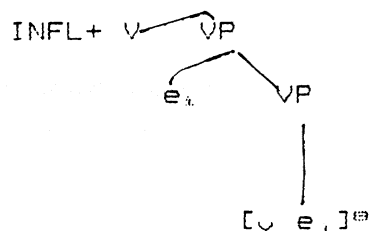
(15a) What do you know how to fix

??(15b) What do you know how they fixed

To explain this difference, we will assume that since non tensed clauses contain dummy inflections, they are considered to be nondistinct from VPs and so do not count as bounding nodes. Notice also that in Romance languages, the verb can move into the INFL slot in the syntax, making the InFL a projection of VP in a sense to be made precise and eventually moving into CP. Therefore, we might claim that in certain cases, even tensed clause IPs do not count as bounding nodes in Romance; allowing sentences like (13).

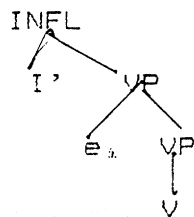
We assume that verbs adjoin to inflection in the syntax in Romance, subsequently moving to Spec CP. If this is the case then adjunction of a WH trace to VP is blocked because the trace of the verb would not be properly governed. The structure is given in (16)

(16)



In languages that do not allow the verb to move to SPEC CP, we assume that the verb moves to INFL in LF. Again adjunction of a WH trace would be disallowed because since the inflection itself is non-lexical, it could not lexically govern the adjoined trace at PF. The structure is given in (17).

(17)

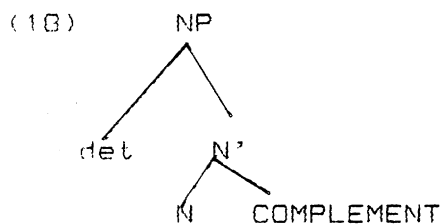


We turn now to blocking illicit movement from syntactic islands. In chapter 4, we considered how to block movement of adjuncts from these constructions so we will concentrate on illicit movement of arguments. We should expect that these cases will be handled straightforwardly given that our analysis is similar to more traditional conceptions of bounding. The only

difference that we must monitor, is that, given that we allow free adjunction, we must insure that the addition of possible landing sites for movement does not permit extraction from islands.

With this in mind, let us consider relative clauses.

We argued in the previous chapter that these categories have the structure given in (18)



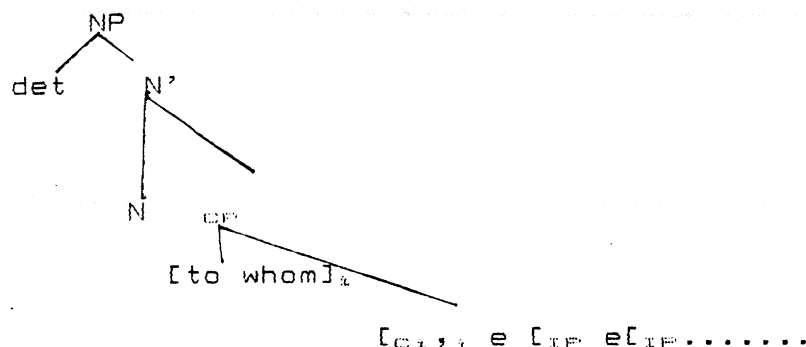
Movement out of an island is blocked in a case like (19):

*(19) [_{CP} What_j do [_{IP} you [_{VP} like [_{NP} the man [_{CP} to whom_i
 [_{IP} Bill [_{VP} wrote e_j e_i]]]]]]]]

Under the standard analysis, there is an NP and CP intervening between the Wh in the Spec CP and the trace inside the relative clause, so sub jacency is violated in (19). Other derivations are also blocked. So for example, if we tried to adjoin to the relative clause's IP, the wh in COMP could transfer its lexical

features to the C and thus the adjoined trace would be properly governed.

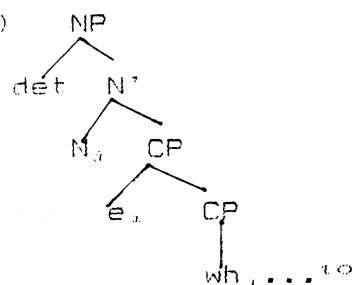
(20)



However, since both CP and NP are still bounding nodes, the NP must still adjoin to CP in order to move from the island."

Relative clauses contain predicates that must be predicated of the head. Predication holds under government between the head and the predicate. Adjunction to CP will block this predication relation. The structure is given in (21).

(21)



Therefore, since all possible adjunction is blocked, extraction from a relative clause is impossible.



The Libraries
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Institute Archives and Special Collections
Room 14N-118
(617) 253-5688

This is the most complete text of the
thesis available. The following page(s)
were not included in the copy of the
thesis deposited in the Institute Archives
by the author:

Page 204

adjunction to CP because, given our definition of subadjacency, direct adjunction to the NP still crosses two bounding nodes; the CP and the NP that is the adjunction site. Once the Wh element is adjoined to the CP though, further movement will leave a trace that is not lexically governed.¹¹

*(23) [_{CP} Who_i do [_{IP} you believe [_{NP} [_{NP} the fact] [_{CP} that [_{IP} John [_{CP} liked e_i]]]]]]]?

Next, let us consider noun complement structures.

Movement from the subject position of a noun complement structure is disallowed. If a noun complement appears in a tensed clause, the IP and NP will count as bounding nodes and the complement must adjoin to the NP in order to obey subadjacency. Further movement will thus leave a non properly governed trace in the NP adjoined position. The structure is given in (24)

*(24) [_{CP} who_i did [_{IP} [_{NP} e_i [_{NP} a picture of e_i]] appear on the table.¹²

WH movement from a complement in object position is allowed even from deeply embedded complements because a complement PP or NP can successive cyclically adjoin to other complement PPs. In certain cases this seems to yield unacceptable results but in others this movement seems perfect. Contrast (25a) and (25b). I do not have an explanation for why 25a is as deviant as it is.

??(25a) This is a man that I read reviews of books about



The Libraries
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Institute Archives and Special Collections
Room 14N-118
(617) 253-5688

This is the most complete text of the
thesis available. The following page(s)
were not included in the copy of the
thesis deposited in the Institute Archives
by the author:

Page 206

Now we turn to "long movement" cases in Romance languages. The basic facts, presented in Rizzi(1982) and Sportiche(1981) show that a WH element can move over one filled CP particularly if that CP governs an infinitival clause. Thus (28) is acceptable while (29) is not.

(28) Non so proprio chi possa avere indovinato a chi
affidero questo incarico.

'I really don't know who might have guessed to whom I will
entrust this task'.

* (29) Questo incarico, che non so proprio chi possa avere
indovinato a chi affidero mi sta creando un sacco di
grattacapi

'This task, that I really don't know who might have guessed
to whom I will entrust, is getting me into trouble.'

In order to rule out (29), Chomsky (1986) makes the following two assumptions:

(30) In English type languages, the lowest tensed clause
becomes an inherent barrier.

(31) Even if only one barrier is crossed per movement, a derivation counts the accumulated set of barriers, and if that number exceeds one, the derivation is ruled out.

While it is clear that these assumptions will rule out the bad cases, they are again, quite peculiar. There is no other operation in the grammar that makes use of the notion "lowest IP". Nonetheless as Chomsky notes, we must specify that only the lowest IP counts, or else we will also rule out a case like (32).

(32) Who do you think [_{CP} e_i that[John said [_{CP} e_i that [_{IP} Mary liked e_i]]].

The notion of "counting up violations" is also not used elsewhere. It is forced here however. Since Chomsky allows adjunction to VP, every movement across a complementizer, will cross only one barrier; i.e the CP that has inherited barrierhood from the IP that contains the embedded VP adjoined trace.

The account that I am proposing allows us to do without these dubious assumptions. In tensed clauses where both IP and CP count as barriers, long movement will be barred as it was under previous approaches to subjacency.

For infinitivals where IP can be non distinct from VP; making intervening IPs not count as bounding nodes, the theory of



The Libraries
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Institute Archives and Special Collections
Room 14N-118
(617) 253-5688

This is the most complete text of the
thesis available. The following page(s)
were not included in the copy of the
thesis deposited in the Institute Archives
by the author:

Page 209

from adjuncts containing infinitival or gerundive clauses is marginal, at worst even in English. Moreover the Barriers theory predicts adjunction to these adjuncts for overt movement because adjunction is only barred to arguments. Consider the following paradigm:

(34) Which monument should we not leave London, without PRO
seeing.

(35) Which monument should we not leave London without taking
Harry to see

?(36) Which monument should we not leave London without Harry
seeing

*(37) Which monument should you not leave London before Harry
sees

Assuming that infinitivals are structures with INFL non distinct from VP, we would expect that we could move directly from the adjunct to the matrix CP, crossing only the matrix IP. A structure is given in (38).

(38) [_{CP} which monument_i should [_{IP} we not [_{VP} leave London]
[_{VP} before [_{IP} PRO visiting e_i]]]¹⁵

By contrast, the tensed inflection heads an IP distinct from VP. Therefore, the category must first adjoin to the adjunct's IP. However, further movement will leave a non-lexically governed trace. The structure is given in (39).

(39) [_{CP} which monument_i should [_{IP} we not [_{leave} London]
[_{IP} before [_{IP} e_i [_{IP} Harry visits e_i]]].¹⁴

If the adjunct is an infinitival, we predict that one could even move across a WH island contained in the adjunct phrase. Here speakers seem to differ. Browning (1987) cites (39) as unacceptable, but a variety of French speakers (personal communication), don't find them deviant. Similar examples, even in English seem not totally unacceptable.

*(40) Quelle personne Jean se demande comment quitter Paris
sans rencontrer

?(41) Which person did John wonder how to sneak away without
leave Paris
seeing.

As we saw in the previous chapter, long movement of an

An anaphor from an adjunct is rigidly excluded by generalized binding.

5.3 Concluding Remarks

In this chapter, we have compared two theories of bounding. We have suggested some empirical problems for the Barriers approach, particularly in regard to parametric variation, the theory of adjunction to noun phrase adjuncts, and movement from noun complements. We have presented a theory where no unmotivated stipulations are made about adjunction sites. However the astute reader will have noticed, that with the exception of noun complement structures, adjunction to categories other than IP seldom leads to acceptability. We will see however, that the ability to adjoin to NP and IP plays a crucial role in determining the distribution of parasitic gaps, a topic that we will take up in chapter 7.

One of the most important differences between this approach and the Barriers approach is this theory's return to an earlier version of bounding theory. Like the theory originally outlined in Chomsky (1973), a substantive class of nodes are selected for bounding. Chomsky (1986) suggests that there is no good reason for taking only NP, IP and CP as bounding nodes. For Chomsky (1986) a substantive theory of bounding nodes gave way to one based on structural considerations. Though all maximal

projections were potential bounding nodes, this property could be removed if a maximal projection happened to be governed.

There is another possible way of dealing with Chomsky's objections; one that tries to give substantive reasons for why just these nodes constitute the inventory of bounding nodes. Clauses and NPs are domains of completion of theta assignment. They are the domains within which "complete functional complexes" are defined. When one looks at subadjacency restrictions from a parsing point of view and thinks of the central problem of parsing as assigning theta roles to categories and categories to theta positions efficiently, then it does not seem at all surprising that the bounding domains should be coextensive with domains in which all of a predicate's theta roles are assigned.

As mentioned, clauses and NPs are the domains within which theta theory operates. So, for example, the theta criterion prohibits the same theta role from being assigned twice within the same theta domain. Consider Noun Phrases for example.

In (42) 'John' is at least three ways ambiguous. It can have the theme reading, the agent reading or the possessor reading:

(42) John's picture

However, as we "add" these theta roles overtly to the NP 'John' can no longer bear these theta roles:

(43) a. John's picture by Rembrandt

b. John's picture of Harry

c. John's picture of Harry by Rembrandt

In (43 a), 'John' can no longer bear the agent theta role. 'John' can be the theme or the possessor but not the agent. In (43b), 'John' can have the agent or the possessor but not the theme role. In (43c) 'John' is limited to the possessor theta role. These facts can all be accounted for if one assumes that the theta criterion applies within NPs to limit theta role assignment. If it does, the same theta role cannot be carried by more than one argument.

Similar reasoning holds within clauses. Thus, one could see (44) as ruled out for the same reason that in (43a) both 'John' and 'Rembrandt' can't both have the agent theta role.

*(44) John painted Bill by Rembrandt

The domains for theta assignment are the NP and the clause. It seems natural, therefore, given the central problem of parsing, that these should also be the domains for bounding.¹⁷

Footnotes

1. Chomsky limits n to at most 1, following the argument given in Chapter 3, where we claimed that since grammars could not employ counting predicates, locality could only be defined by employing notions like adjacency.

2. Henceforth, we will refer to this condition as the CED.

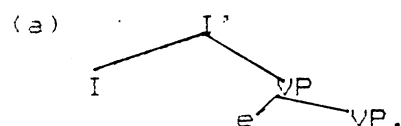
3. We know that this functions like a standard A-bar operator because heavy NP shift licenses a parasitic gap.

4. Moreover, we can delete the heavy NP adjoined trace before LF and so the heavy noun phrase can function as an unambiguous binder for the trace left in its theta position.

5. We will see below that if a bridge verb voided the bounding status of a CP for everything inside its complement, we would incorrectly allow movement across a series of previously filled CPs as in a case like (a).

*(a) Who did you wonder who knew who saw.

6. We could also derive this sentence by first adjoining the object to the embedded VP. We will assume that in English, the verb does not move into the inflection position, and so the resulting structure of this movement would be (a), where the adjoined trace would not be properly governed.



We will discuss adjunction of the verb in languages that allow movement of V to Infl below.

7. These sentences are from Mouchaweh(1985) and Rizzi(1982).

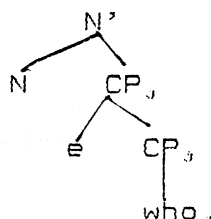
8. As we will see below, we must block adjunction to VP in order to block super long movement.

9. Direct adjunction to the NP is blocked because, given that our definition of subjacency is not stated in terms of exclusion, the trace inside the relative clause will still be separated from the NP -adjoined trace by two bounding nodes.

(a) [NP wh_i [NP [CP to whom [IP e_i...]

10. See Williams(1980) and Hornstein(submitted), who justify the assumption that predication is only under government. One might think that the predication index could percolate from the operator to the adjoined CP, thus allowing the predicate to meet the government restriction as in (a)

(a)



Evidence that predication can only hold between the CPs that immediately dominate predicate operators comes from consideration of the following contrast. Some speakers can adjoin adverbs to S' even in embedded constructions (assuming topicalisation intonation, and controlling for the well known parsing strategy that preferentially interprets adverbs as right adjoined structures (see Kimball(1973), Fodor and Frazier(1978)).

(b) Yesterday, how much money did you have?

(c) I want to know right away, yesterday, how much money you had.

However, even for these speakers, an adverb cannot be adjoined to an S' in a predication structure.

*(d) I know a man, yesterday, who had some money

Notice that both this theory and the Barriers model will have to make the assumption that adjunction to the CP relative clause is forbidden. As discussed above, Chomsky must block adjunction to CP in relative clauses to block antecedent government into and movement out of, these structures. Since he assumes that relative clauses are not theta marked by the verb, he cannot use the prohibition against adjoining to arguments in this case. Therefore, he must independently block adjunction to relative clauses.

11. It has been noted that movement out of infinitival complements yields improved results as shown by (a).

?(a) Who_i did you accept [NP a proposal [CP PRO to hire e_i]]?

Notice independently though that only tensed clauses can appear in appositive structures in English.

- (b) John, who anyone likes, is in the room.
 *(c) John, to like, is in the room.

Therefore it seems reasonable to expect that infinitival complements will appear as direct daughters of N' without an intervening CP. Notice that the PRO in this case gets a controlled reading in this structure. If we assume that obligatory control is a property of governed PRO, as argued by Bouchard (1984), Sportiche (1983), and Hornstein & Lightfoot (1987) then the structure of these infinitivals must be (d).

- (d) [_{NP} a [_{NP} proposal [_{IP} PRO to hire e_i]]]

We could then handle this case by assuming that, since in English infinitivals are non distinct from V, the IP would not count as a bounding node. In this case, no adjunction to IP would be required.

As we will see below, the Barriers framework as stated, does not distinguish between tensed and infinitival noun complements, so some additional assumption would be needed in this theory as well.

12. We will discuss movement from the subject position of an infinitival or small clause subject in Chapter 6.

Even if IP did not count as a bounding node movement from subject position will be ruled out in certain cases. For example, movement from the determiner position of a matrix sentence will be barred at least in a language with overt movement and the doubly filled COMP filter that we discussed in the previous chapter because the determiner position could not be properly governed without violating this filter. Thus a case like (a) is correctly excluded.

- *(a) Combien_i sont [_{IP} e de bons bons] sur la table?
 How many are candies on the table
 How much candy is on the table.

13. Notice that, in order for 25b to respect the condition of lexical proper government, we must assume that nouns as well as verbs can govern objects of prepositions that they govern. This is a marked option that English and some Germanic languages (as opposed to Romance languages) allow.

14. Independent of this question, we must assume that extraposition applies in PF because we have argued that predication is a head - head relation between an empty operator and a head that applies under government. Extraposition will move the empty operator from the position



The Libraries
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Institute Archives and Special Collections
Room 14N-118
(617) 253-5688

This is the most complete text of the thesis available. The following page(s) were not correctly copied in the copy of the thesis deposited in the Institute Archives by the author:

← Page 218

A Comparison of the Barriers and the WAHL Frameworks

In the previous chapter we presented two different approaches to the theory of bounding and the ECP. The first approach, that of Chomsky(1986) argues that both ECP and bounding are subject to a unique set of restrictions that specify licit government relations and can be captured as a uniform set of barriers. In the Wahl framework, a connection between bounding and the ECP is also established but the connection comes from the fact that every trace left by overt movement will be subject to the PF ECP. Since the notion of L-marking from Barriers and the notion of lexical government are rather similar, it is difficult to tease apart different predictions made by these theories, although some advantages of the WAHL approach were discussed.

A more direct empirical test of the conceptual organization of the two frameworks comes from looking at the relationship between bounding phenomena and the part of the ECP that governs LF phenomena (antecedent government or antecedent government/lexical government in the Barriers framework and Generalized Binding in the WAHL approach). The Barriers approach makes the clear prediction that we should be able to find a unique set of barriers governing both overt and LF movement. The WAHL theory, on the other hand claims that any restrictions that apply to both types of movement should stem from the Binding Theory that applies to the output of both processes at LF. Since the domains set up by barriers and binding theory are different,

these theories will rule out illicit movement in these components differently. In this section, I will present some arguments for claiming that LF movement results in syntactic structures of unrestricted quantification. After presenting empirical support for this idea, I will try to show that it is fully consistent with the WAHL framework, but undermines the link between LF ECP effects and bounding conditions inherent in Barriers.

6.1 Superiority

The first case that I would like to consider is the so called "Superiority Condition". As in a variety of LF movement structures, superiority examples exhibit what can loosely be called a subject/object asymmetry in the sense that moving an object across a subject at LF results in unacceptability as shown by (1)⁴

- (1) a. *What did who buy
- b. *John wonders what who bought
- c. Who bought what?

Within a Generalized Binding approach such sentences are treated as Binding Theory violations.

- (2) [_{CP} [_{NP3} who_i] [_{NP3} what_j] [_{C'} [_{CP} [_{NP} t_i buy t_j]]]

In (2) ' t_1 ' has AGR as an accessible SUBJECT. Therefore ' t_1 ' must be bound in the matrix clause. In English, a WH element can only move by WH- raising to a COMP that has a +WH specifier.³ As discussed in the previous chapter, we assume that while the NP containing the LF adjoined trace c-commands the rest of the sentence, the WH element 'who' that is adjoined to the SPEC CP position does not c-command and so does not bind the subject trace t_1 at LF, as required by GEBI.⁴ A similar account extends to (1b).

(1c) is acceptable because in its LF structure, the subject trace ' t_1 ' is bound, either because the subject moves into the complementiser at s-structure, or because, since the object trace does not require a binder for reasons discussed in the previous chapter, the WH element does not have to c-command its trace in order for the trace to be cinterpreted with it. This is shown in (3)

(3) [_S [_{COMP} t_1 [_{what} t_2 [_{who} t_3 [_C [₁ t_1 AGR buy t_2]]]]]

An important fact about superiority effects is that they are cancelled in certain unexpected cases.⁵

(4) a. Which book did which boy buy

b. John wondered which book which boy bought

A recent analysis by Pesetsky (1985) tries to explain the difference between the examples in (1) and (4) by suggesting that the latter have structures like (5) in LF.

(5) [[_{COMP} Which book, [Which boy AGR buy t_i]]

Pesetsky's analysis claims that in (5) which boy is in situ in LF. Therefore there is no gap. Pesetsky uses a discourse principle to decide which WH elements are moved at LF and which remain in situ. Elements that are discourse linked (hence "d-linked") in the sense that they refer back to some previously mentioned element in the discourse remain in situ and are linked to a Q(uestion) operator in the COMP position at LF. WH elements that introduce new elements into a discourse (henceforth "non d-linked" categories) are interpreted by WH-R.⁶ The absence of a trace suffices to explain the absence of superiority effects. Before arguing that Pesetsky's elegant suggestion is nonetheless incorrect, I will present an alternative.⁷ As argued in Reinhart(1983) and Aoun and Hornstein(1985), I assume that an NP may be c-commanded by a containing NP. I will also assume that Wh movement in LF does not pied pipe the category that contains a Wh element along with that category. With this in mind, consider (6) which has the LF representation (7).

(6) What book did whose mother buy.

(7) [_S[_{COMP} What book_j whose_i] [_S[_{NP} t_i's mother] AGR_{EC} buy
t_j]]

(6) will be well formed if we take 't_i' to be an A/C anaphor which would then have no accessible SUBJECT in (7). NP_{EC} is not an accessible SUBJECT because NP_{EC} c-commands t_i given our definition of c-command. Thus coindexing t_i with NP_{EC} would yield a Principle C violation and our definition of accessibility disallows a subject from being accessible for a category if Principle C would be violated. AGR is not an accessible SUBJECT because indexing AGR and 't_i' would yield a principle C violation given the required coindexation of AGR and NP_{EC}. Consequently, 't_i' has no domain and the Generalized Binding theory does not require it to be bound.

We can extend this account to (4) if we assume that the two sentences have structures similar to (7) at LF.¹⁰

(7) b. [_S[_{COMP} which book_j which_i] [_S[_{NP} t_i boy] AGR buy
t_j]]

Since the structures (7a) and (7b) are identical, our account of (5)'s acceptability suffices for (4) as well.

In order for this approach to be viable, it must be the case that QR need not move a whole NP but can move quantifier phrases alone at LF. This makes the LF representation that is submitted to the Binding theory look like standard versions of the predicate calculus that allow unrestricted quantification. We should now discuss the interpretation of these structures. Two possibilities present themselves. One possibility would be to interpret these structures as structures of restricted quantification. This proposal is in the spirit of one proposed in Williams (1986) who notes that none of the information needed to interpret structures like (4) restrictively, is lost in a representation like (7). Principles of interpretation for these structures are given in (8) and the output interpretation is given in (9).

- (8) a. The QP in an A' position in LF is interpreted as the quantifier. The quantifier can be given a variable index from the variable to which it is bound.
- b. The N' governed by the quantifier or a trace with which it forms a chain is the restriction.
- c. The trace of the QP (or the NP containing the QP trace) is the variable.
- d. The scope domain of the quantifier is the IP or CP that is in construction with the QP.

(9) Wh x: wh y [_{IP} [_{NP} y [_{N'} boy]] AGR buy [_{NP} x [_{NP} book]]

Alternatively, we could treat these structures exactly as they are treated in the predicate calculus, i.e. unrestrictively, with the normal translation procedures into first order structures that apply to universal and existential quantification. In this case (7) would receive the interpretive structure (9b)

(9b) Wh x and Wh y: x a book y a boy and y bought x.

We know that such translation procedures exist for wh, existential, and universal quantification and so either procedure will give the right interpretations for these cases.⁹ In the appendix at the end of this chapter we present an explicit algorithm that translates our LF representations into expressions fully appropriate for semantic interpretation.

Whatever the merits of this more inclusive claim however I would like to show that my proposal for the treatment of superiority is to be preferred to the one proposed in Pesetsky (1985). First, adopting Pesetsky's suggestion will require adding an additional rule of interpretation. Sentences such as 'who bought what' are interpreted as general questions. They require buyer/buyee pairs as appropriate answers. This interpretive requirement has been standardly related to the fact that at LF such sentences have two operators in COMP.¹⁰ If one adopts the approach advocated in Pesetsky (1985) we will have to

add a second disjunct to this interpretation procedure. The reason is that 'Which book did which boy buy' is also interpreted as a general question. But in this case, according to Pesetsky (1985), which boy has not moved to COMP in LF. Nonetheless to derive the correct interpretation, it must be associated with the COMP filled by which book. This end can be achieved by making the rule of interpretation disjunctive: interpret 'Wh....wh... wh...' as a general question just in case at LF all the wh-operators are in the same COMP or they are in situ but are somehow "associated" with the COMP. The main point is that Pesetsky's (1985) account forces wh elements to be interpretively linked by two very different LF procedures.

The same complication does not affect our proposal. In both 'who bought what' and 'which book did which boy read' there are two wh elements in COMP at LF. Consequently the interpretation rule for multiple questions need not be revised to extend to these cases.

An empirical advantage of our proposal is that it extends to cases of complex wh phrases which are not D-linked.

(10) What type of book does what type of man read ?

(10) uses what as a specifier. In contrast to the examples cited above, (10) involves no presupposed set of appropriate answers

among which the answerer must select. Interpretively, what in (10) is a non D-linked wh-operator. Nonetheless, (10) is as acceptable as (6) and (7). This clearly constitutes a problem for the analysis in Pesetsky (1985) which predicts that since the quantifier is non d-linked and therefore must move to its scope position, the structure in (10) should be on a par with (1) and (2). As what is non D-linked it must move at LF. We should observe the standard superiority effects.

By contrast, our approach predicts the acceptability of these sentences. In (10) what is in specifier position. At LF (8) has the structure (11):

(11) [_{Comp} t_i] [what type of book_i, what_j] [_S [_{NP} t_j, type of man] read t_i]]

(11) is virtually identical to (6) and the explanation we advanced for the well formedness of (6) will carry over wholesale to this case. In short, lack of superiority effects seems tied to the fact that in cases such as (6), (7) and (11) we are extracting a wh-element out of a more complex NP while in cases such as (1a,b) this is not a possible option. Whether the Wh-operator is D-linked or not seems to be irrelevant.

Last of all, Pesetsky's analysis has little to say about the acceptability of cases like (12)

- (12) a. What did whose mother buy
 b. What did which mother buy

Whose is clearly not a D-linked operator, yet (12a) is as acceptable as (12b). Pesetsky's analysis could be extended to cover cases such as (12a). Thus it might be proposed that whose mother remains in situ in LF. However, this would sever the connection between D-linking and LF movement which lies at the heart of the account. Another possibility would be to countenance movement of whose in cases such as (12a) to yield LF structures such as (11). This would, in effect, adopt our proposal for cases such as (12a) while retaining the D-linking/non movement hypothesis for (12b). However, this would be an odd conclusion since accepting our analysis for (12a) would render Pesetsky's analysis of (12b) superfluous.¹¹

We conclude that the approach to superiority discussed in this chapter is preferable to the one proposed in Pesetsky (1985). The analysis is simpler. It dispenses with an additional interpretive rule. It is more empirically adequate.¹² It correctly predicts that complex NPs with non D-linked Wh specifiers will act just like D-linked ones as regards superiority effects. It also has greater empirical coverage.

It covers cases that the D-linked/non movement account does not apply to.¹³

6.2 Pronouns as Bound Variables Revisited

Again, this rather simple emendation to the QR rule allows us to simplify the theory of scope assignment in rather striking ways.¹⁴

The examples that we have discussed so far indicate that QR can move the QP alone and leave the rest of the quantified NP in place. There is some evidence that QR can only move a QP and must leave the containing NP in situ. As Williams (1986) observes, 'Move' in LF, unlike 'Move' in the syntax cannot alter binding possibilities. Consider cases such as (13):

- (13) a. Which picture of himself_i did Helen persuade
Frank_i that Sheila liked
b. *Who persuaded Frank_i that Sheila liked which
picture of himself_i

(13a) is acceptable while (13b) is not. The acceptability of (13a) can be accounted for if we assume a theory of Reconstruction like the one outlined in Hornstein (1984). On this sort of theory an NP need not actually be put back into a pre S-structure position. Rather, we can adopt an interpretive

principle that states that an NP can be interpreted from any of its trace positions. If principle A is met from any of these positions then the indicated reading is permitted. (13a) has (14) as its LF phrase marker.

(14) [which_i [t_i picture of himself]_j [Helen persuade Frank [t'_j that [Sheila liked t''_j]]]]

Observe that given this theory of Reconstruction, NP_j, the noun phrase which contains 'himself', can be interpreted from position t'_j. From this position Frank is a permissible binder and the indicated reading of (13a) is accounted for.

What is important, however, is that a similar process is not available for (13b). To form a multiple question the WH-in-situ must move to the matrix wh filled COMP. If we assume that LF movement moves an entire QP, then the LF interpretation for (13b) is (15).

(15) [Who_i, [which picture of himself]_j [t_i persuaded Frank [t'_j that [Sheila liked t''_j]]]]

If this is correct though, we predict that we should be able to bind himself to Frank. by interpreting NP_j from position t'_j, just as we did with NP_j in (13). Or, to put this another way, we

are left with the question of why wh-raising in LF cannot change binding possibilities in the way that overt movement can.

If we assume that LF operations only move operators, ie that 'Move' in LF can only move QPs, then we have a straightforward explanation for the unacceptability of (13b). At LF (13b) has the structure (16).

(16) [_{wh}_i which_j [_{t_i} persuaded Frank [_{S_i} [_{t_j} that [_{S_i} [Sheila like [_{t_j} picture of himself]]]]]]

In (16) himself has S_i as its domain. However, it is not bound in S_i, so principle A is violated and the unacceptability of (13b) is accounted for.

In short, we have evidence that in LF we move operators and only operators to adjunct position.¹⁵

A last feature of this proposal is that it makes the correct predictions in regard to Reconstruction phenomena. As mentioned above, we follow Hornstein (1984) and assume that Reconstruction is an interpretive rule which permits an element indexed with a trace to be interpreted from that trace position.

In (17) Reconstruction permits his to be bound by everyone because the wh-phrase is indexed with t_j, and so it can be

interpreted from this position. In this position everyone would c-command his and bind it. Hence through the offices of the LF interpretive rule of Reconstruction, pronoun binding of his by everyone is permitted in (17).

- (17) [[Which picture of his_i mother] $_j$ does [everyone $_i$ love t_j]]]

Williams (1986) reiterates an apparent problem first discussed by Higginbotham (1980) for this view of reconstruction. Consider a sentence such as (18).

- (18) Which picture of which man did he like

Here he cannot be bound by which man. Given our theory of reconstruction, (18) has the structure (19) at LF.

- (19) [Which $_i$ which $_j$ [_{NP₁} t_i picture of t_j man] [he $_i$ like t_{1i}]]]

Note both Wh-operators are in adjunct position. Observe that if he were bound by which $_j$, then with NP₁ interpreted from position t_i (via Reconstruction) we would have a principle C violation.

If we assume that R expressions must always be locally A-free in any position from which they can be interpreted then (19) is predicted to be ill formed under the indicated interpretation.

Observe in (19), that from position t_k the variable is locally A-bound and so the structure is out.

Observe that we've assumed that which_k and which_j have been moved out of the NP that contained them. This follows if we assume that operators must be in adjunct position in LF if they are to be interpreted as operators. At S-structure, prior to LF movement, they are not in adjunct position. Rather the phrase containing them is. In other words at S-structure these wh-operators are no more in adjunct position than who₂ is in (20).

(20) Who left after Bill met who₂.

In (20) who₂ is in an adjunct, (viz, after Bill met who₂), but it must still move at LF. Similarly for the which operators in (19). Though they are both contained in an NP in COMP they themselves are not in adjunct position and so must move. This results in a structure like (18).

Assuming a theory that generates structures where only the QP is moved has more far reaching consequences than the empirical ones that I have outlined up to this point. More importantly this analysis allows us to employ a uniform algorithm for scope interpretation at LF and thus allows the theory to sidestep some of the criticisms levelled against it in Williams (1986). Williams, contrasting cases like (13) and (19) correctly notes

that the standard theory for LF interpretation must allow LF structures to be reconstructed in an arbitrary manner. Structures like (18) force us to assume that scope assignment can be based on either the S- or LF structures, but structures like (17) must be interpreted with the LF-moved elements in their d-structure positions. Williams notes that this removes much of the explanatory power from a theory that assumes LF. For arbitrary derivations, the claim that correct scope assignment or pronoun binding results from applying independently motivated syntactic principles to a new level of interpretation is negated and interpretation proceeds as if this level did not exist. Given the theory discussed in this chapter, we allow a uniform rule of scope interpretation. Cases like (18) are not counterexamples to this procedure because since we do not pied pipe in LF, reconstruction does not even apply in these cases. We deal with reconstruction of syntactically moved WH elements by assuming the reconstruction principles first argued for in Hornstein (1984). It is important to note that we interpret the results of LF and S-structure movement from a unified LF structure. Since Hornstein's principle allows LF interpretation to take place from any trace position we are not faced with the problem of disregarding outputs of the LF component in unprincipled and ad-hoc ways. As in many other cases, the syntactic and LF components will overgenerate possible interpretations that will be filtered out by interpretive principles.

6.3 General Theoretical Consequences of the Account

In the last section, we argued that QR crucially yielded structures of non-restrictive quantification. This analysis in turn bears on the more general question of the proper theoretical formulation of the ECP and on the underlying assumptions of the Barriers framework of Chomsky (1986).

6.3.1 Movement from NPs

Allowing QR to simply move a determiner from a noun phrase is equivalent to allowing violations of the Left Branch Condition at LF. However, it is well known that such violations are not possible (at least from subject position) at S-structure, as shown by the contrast in (21):

(21) a. Which pictures_i do you believe _{TP}[_{NP}[e_i] will be in the exhibition]

*b. Which_i do you believe _{TP}[_{NP}[e_i pictures] will be in the exhibition.

One may try to rule out (21b) by adopting any of the many versions of the ECP. Some authors try to reduce all ECP cases to instances of lexical government. Chomsky's (1981) original version allowed a structure to satisfy the ECP if any empty category was either lexically or antecedent governed.¹⁴ Later approaches (including the Barriers approach of Chomsky

(1986) try to reduce all cases to antecedent government. Let us consider all of these approaches in turn.

(21b) violates lexical government under a variety of assumptions. We could assume, as in Chomsky (1986), that COMP is not a lexical item and thus there is no category of the appropriate type to govern the trace of 'which'. Even assuming the COMP is lexical, at least one maximal projection (the NP in subject position) and arguably two (the S or IP dominating the subject NP) intervenes between COMP and the trace. Therefore, assuming standard theories of government, the trace in (21b) does not appear in a position from which the COMP can properly govern it. Note that structures like (22), generated by our analysis at LF, also violate this notion of lexical government.

(22) a. Which book did which boy read.

b. Which book_i which_j did_i [_{NP}[_{NP}e_j boy] read e_i.

Given standard versions of antecedent government, antecedent government will be blocked under similar assumptions. The subject NP (a maximal projection) blocks government so even though 'which' as a coindexed WH phrase is a potential antecedent, it cannot antecedent govern inside the subject NP. Given these assumptions, neither disjunct of the ECP is satisfied in (21b) and so the structure is out under any of the versions of the ECP outlined above.

Notice however, that if the above versions of antecedent and lexical government apply at LF, then structures like (22) above, where we have simply moved a determiner out of subject position will be ruled out as well. Therefore, in order to allow structures like (22), we must either reformulate antecedent government in terms of Generalized Binding as suggested above or loosen the locality conditions on antecedent government or lexical government within a more traditional ECP framework.

We can rule out such structures at S-structure but allow them in LF in one of two ways. We might adopt the organisation of the grammar proposed WAHL (forthcoming) and discussed in the previous chapter. Recall that WAHL proposes that lexical government applies at PF, subject to conditions similar to those that we mentioned in our discussion of lexical government above. Since syntactic, but not LF movement feeds PF, we predict that (21b) but not (22) would be ruled out by this condition. (21b) and (22b) satisfy Generalised Binding conditions which apply at the LF level. WAHL requires that empty categories must satisfy all the locality conditions that apply to them. Therefore even though (21b) is licit at LF, it is illicit at PF and thus the structure is ruled out.

A second alternative is provided by the Barriers framework of Chomsky (1986), where structures like (21b) would be ruled out

by the Subjacency Condition. As discussed in the previous chapter, every category that is not directly theta-marked by a lexical category counts as a barrier for movement in this system. Thus the subject NP in (21b) is a barrier. In addition, barrierhood can be inherited by a maximal projection that dominates other barriers. Therefore, movement from the determiner position of a subject NP crosses two barriers, which is disallowed by bounding theory. Thus (21b) is out as a subjacency violation. Since subjacency applies at S-structure, structures like (22b) which are generated at a post S-structure level are not ruled out by this condition.

Chomsky can allow in these structures at LF by using a reformulated definition of barrierhood that interacts with the theory of antecedent government, though crucially not with the theory of movement, to include the so-called "minimality" condition. This condition is given in (23)

(23) In the structure, a [₀...b...]

a projection

"a does not govern b if g is...the immediate projection of g, a zero level category distinct from b" (that excludes a (ASW))¹⁷

Since determiners of NPs are not separated from antecedents in COMP by the "immediate projection" (N^0) of the NP, the NP does not count as a barrier for them. Thus, under this definition, the DPs in 26b would antecedent govern their traces because the

NP dominating the determiner position would not count as a barrier and so the IP dominating the NP which is not an inherent barrier in Chomsky's system, would also not inherit barrierhood. Thus no barriers would separate the QP from its trace.

Notice however that we are obtaining this result at the cost of allowing barrierhood to be determined in two separate ways. If a category is dominated by an ungoverned maximal projection it cannot move from that maximal projection in the syntax. A category can move from an ungoverned maximal projection in LF, however if it is not dominated by the first lexical projection of that category. Chomsky tries to justify this distinction by claiming that the minimality condition enforces a uniqueness of government, allowing a category to be governed only by its closest possible governor. While this condition does establish a unique governor for the complement of a noun phrase, it does not serve this function for the determiner as shown by a case like (24), where even though the N governs the determiner position, and is the closest possible governor, the minimality condition allows an element that appears outside this NP to govern this position.¹⁹

(24) tell [_{NP} North's [_N contact [in the department]]

Thus it seems that the minimality condition does not have an independent conceptual motivation and is inconsistent with the

any of the categories that intervene between the CP and determiner trace is crucially disallowed within the Barriers framework. The PP inside the complex NP is theta marked by the head of the category which removes it as an adjunction site under the assumption that theta marking cannot apply to part of an adjoined structure.²⁰ The NP, also being theta marked, is not a possible adjunction site. Since it is also not an L-marked category, it is a barrier and the immediately dominating IP inherits barrierhood. Therefore antecedent government of the determiner trace by the WH in matrix CP position is doubly blocked by the intervening NP and IP barriers.²¹

One could reconcile structures like (24) if one allowed optional pied piping at LF. Pied piping the entire noun phrase as in (27) yields a structure that satisfies the Barriers definition of antecedent government.

- (27) [pictures about what]_i [who]_j [_{IP} e_i [said [that
[_{NP} [e_j] were on sale.]]]]]

Pied piping the PP yields a structure that satisfies lexical government.

- (28) [[about what]_i [who]_j] [_{IP} e_i said that [_{NP} [pictures
[e_j]_{NP}] were on sale.]]]

The problem with this strategy, is that in order to explain the contrast between (13a&b) repeated as (29), we have had to assume that pied piping was prohibited in LF.

(29) a. [Which picture of himself_i]_j did Helen persuade Frank_i that Sheila liked e_j.

b. *Who persuaded Frank_i that Sheila liked which picture of himself_i

In other words the Barriers framework would force us to permit or prohibit pied piping in LF on an ad hoc basis. The account in terms of Generalized Binding does not depend on LF pied piping and so does not face the same problems.²²²

6.3.2 Movement from NP in the Romance Languages

The last section sought to show that the locality restrictions on movement from NP were better handled given the locality conditions imposed by generalised binding, rather than those of a theory of antecedent government that was subject to more or less the same conditions as movement. In this section, we will reinforce this conclusion by consideration of some facts about movement from NP in Romance.

As discussed in the preceding section, Chomsky(1986) rules out all overt movement from the subject position of a matrix clause because the NP, not being L-marked is a barrier, and the IP inherits barrierhood from the NP. (30) from French and Italian shows that this claim is too strong.

(30) C'est [NP un homme [CP dont_i [IP [NP le frere [e_i]]

This is a man of whom the brother

vent devinir peintre]]].

wants to become a painter

This is a man whose brother wants to become a painter

?(31) Giorgio, [CP [di cui_i] [IP [NP l'onesta [PP e_i] e credo,

of whom the honesty I believe

notta a tutti²³

is known to everyone

We adopt Aoun(1985)'s analysis of NPs in Romance for these structures. Aoun claims that the determiner is in an A-bar position in these languages and can thus count as an accessible subject for the PPs inside the phrase. Following both Sportiche(1979) and Cinque(1980), Aoun claims that only NPs that can be coindexed with the determiner position can be fronted from inside an NP in either subject or object position. Sportiche claims that in French,

Chomsky's treatment of overt extraction from the subject position also runs into problems when we consider the case of extraction from the subject of a small clause in French.²⁷

In French, (33a) can be formed from (33b) by moving the quantified expression from the object NP into an adverbial position inside the verb phrase.²⁸

(33) J'ai [_{NP} beaucoup_i mange [_{NP} e_i [_{NP} de gateaux]].

(34) J'ai mange [beaucoup [de gateaux]].

I ate a lot of cakes.

The trace of the quantifier inside the determiner cannot be lexically properly governed because the verb 'manger' does not theta mark it. Therefore, it must be antecedent governed. In order for this to happen, it must adjoin to the V⁰ projection of the matrix verb or else the verb will "protect" the determiner position from government by minimality.²⁹

Problems arise though, when we consider further cases. (35) shows us that quantified NPs can appear as subjects of small clauses. (36) shows however, that these quantifiers cannot move from subject position.

(35) [J'ai entendu [_{NP} [_{NP} beaucoup [d'etudiants] raconter des mensonges]].

I heard a lot of students telling lies.

??(36) J'ai [_{VP} beaucoup_i entendu [_{IP} [_{NP} e_i d'étudiants] raconter des mensonges]].

The unacceptability of quantifier movement from this position is unexpected because in this case the matrix verb will L-mark the its complement IP and the specifier of IP will also be L-marked by SPEC-HEAD agreement. Therefore no barriers for movement are crossed and so the structure is licit by the bounding theory.

In addition, the assumptions that we made to allow 'beaucoup' to antecedent govern from its V^o or VP adjoined position, should also allow it to antecedent govern in (36). The problem is that there is no way to rule out (36) without also ruling out (43).

Moreover, we cannot rule out (36) by bounding theory, or we would predict that (37), which is perfectly acceptable, would also be deviant.

(37) Combien les guerres ont-elles laisse d'hommes angoisses

How many the wars did they leave men anguished

How many men did the wars leave anguished?

(37b) [_{CP} Combien_i [_{IP} les guerres ont-elles [_{VP} e_i [_{VP} [_{VP} laisse [_{IP} [_{NP} e_i d'hommes] angoisses]]]]]].

Unfortunately, this analysis would also allow (38)

*(38a) Quels les guerres ont-elles laisse d'hommes anguisses
 which the wars did they leave men anguished

Which men did the wars leave anguished?

(38b) [_{CP} Quels_i [_{IP} les guerres ont-elles [_{CP} e_i [_{VP} laisse [_{IP}
 [_{NP} e_i [d'hommes] anguisses]]]]].

We will now try to account for this range of data in terms of the generalised binding framework, now crucially supplemented with principles of lexical government for overt movement. Let us first consider the case of 'beaucoup'. Following the typology of quantifiers motivated in Aoun and Hornstein(1985), 'beaucoup' is an A'-anaphor; a purely non-referential category. Thus it must be bound in the domain of its first accessible subject. In a case like (35), the first accessible subject is the subject of the embedded clause, where the category is bound and so the sentence is good.

Movement out of a small clause as in (36) is barred however, because the quantifier must be bound in the domain of its first accessible subject, which in this case is the NP subject that contains it. Even if the category adjoins to the NP, it will leave a trace that is not bound in the domain of its

first accessible subject, and so the structure is ruled out by generalised binding.

(39) J'ai beaucoup_i [VP entendu [IP [NP e_i [NP e_i
[d'étudiants]] raconter des mensonges.

"Combien", which is a WH quantifier and an A/C anaphor has no accessible subject in the small clause. It cannot be bound to the NP that contains it because this would yield a Principle C violation. Therefore it can move to the matrix clause in one swoop leaving a trace that is lexically governed as shown in (40)

(40) [CP combien_i [les guerres ont -elles laisse [IP [NP e_i
[NP d'hommes] angoisses.]]

We rule out movement of 'quel' by claiming that it appears in a non-adjoined determiner structure. Since it, like 'beaucoup', is an A/C anaphor, it can move from the determiner position in one swoop, but this movement leaves behind a non-properly governed trace as shown in (41a). Adjunction to NP will cause failure of lexical government as shown in (41b)

(41a) [CP quels [IP les guerres ont elles laisse [IP [NP
[e]] d'hommes angoisses.

- (41b) [CP quels [IP les guerres ont-elles laissees [IP [NP e_i
[NP [e_i] d'hommes angoissees.

It is interesting to note that movement of 'beaucoup' is acceptable when it accompanies movement of the head of the phrase as shown in (42).

- (42) J'en_i ai beaucoup, laisse [IP [NP_i e_j [NP e_i]] faire
des

I-them have many let make
betises
stupidities.³¹

I let many of them do stupid things.

Given our idea that in order to be a subject for a category, an NP must contain all of its obligatory parts, we would claim that this movement is allowed because as the subject NP_i is missing its head, it is not a subject for the determiner.

6.4 Conclusion

Like the Barriers approach, the Generalized Binding framework also seeks to show that "antecedent government" does

not add its own set of locality restrictions to universal grammar. Rather than reducing the locality conditions to those needed independently for bounding theory, Generalized Binding replaces antecedent government with a theory of "generalized" binding, subject to locality conditions needed independently to govern the distribution of the set of A-anaphors. The analysis presented here suggests that "antecedent government" should in fact be governed by restrictions on binding rather than bounding theory.

Finally, the approach provides important support for the theory developed in this thesis and in WAHL (forthcoming) because the analyses presupposed by it for QR and WH-R are so naturally accommodated into this model.

Footnotes

1. Cases of superiority that do not seem to involve subjects will be considered below.

2. Cf. Chomsky (1973) for the first extensive discussion of these data.

3. This was first pointed out in Aoun, Hornstein, and Sportiche (1981). This generalisation also seems true of all languages that allow extensive WH-movement in the syntax. For example, although French allows movement to a non-explicitly WH-marked matrix clause, this type of movement is barred in embedded positions as shown by the contrast between (a) and (b).

(a) *Je sais que tu vois qui. (with the indirect question reading) I know who you saw

(b) Je sais quoi faire.
I know what to do.

4. For details cf. Aoun (1985), WAHL (forthcoming) The important point is that COMP is the only possible binder but it cannot bear the index 'i' if what_i was in COMP before

who_i. As what_j is in COMP prior to who_i the sentence is unacceptable.

5. This fact was first observed by R. Kayne.

6. A d-linked phrase appears in a discourse like (a) below and is peculiar if used to open a discourse. Non d-linked elements can be the initial elements of discourses like (b):

(a) John, Fred, and Harry walked into the room.

Which man seemed to be the happiest?

(b) Who do you think is the happiest person in the world?

7. A version of this alternative was worked out in collaboration with N. Hornstein and appears in the proceedings of NELS 17.

8. Cf. Higginbotham (1983) and (1985) for some discussion of the parallel role an NP subject and determiners play in noun phrases. Both are able to "saturate" the argument position nouns inherently have. The principal difference is that NP subjects introduce an additional conjunct into the interpretation of the NP, (cf. Higginbotham 1983: 416):

(i) the book = [the x: book (x)]

(ii) John's book = [the x: (book (x) & R (x, John))]

9. Our analogous treatment of the LF of whose and which can be seen as extending to LF operations the interpretive similarities Higginbotham described.

Our approach also handles the differential acceptability of (a) and (b).

John wondered

- a. Who said (that) whose pictures were on sale
- b. *Who said (that) who left the party

The structures for these cases at LF would be (c) and (d):

(c) [[_i Who_i whose_j [[e_j pictures] were on sale

(d) [[_i Who_i who_j] [e_i said that e_j AGR left the party

The LF for (c) is analogous to that of (6) above and t_j has no accessible subject as it cannot be bound to the NP containing it without inducing a Principle C violation, nor with the Agreement marker that is coindexed with its containing NP. (d) has a structure like (2), and as in that case the Agreement marker acts as an accessible subject but the trace e_i is not bound in this domain and thus a Principle

A violation results.

10. Cf. Chomsky (1973), Higginbotham and May (1981), Aoun, Hornstein and Sportiche (1981).

11. Pesetsky briefly considers linking superiority effects to a specifier, non specifier distinction; interpreting this distinction within an analysis proposed by May and Gueron (1983). He suggests that the unacceptability of a case like (a) is problematic for this approach:

(a) I need to know who(m) how many people voted for.

*(a.1) I need to know [who(m)], [how many people], [e_i voted

S

for e_i,

Notice that this effect is directly related to the type of quantifier that occurs in the specifier position as shown by the acceptability of (b)

(b) I need to know who(m) what type of people voted for.

which candidates

(b.1) I need to know [who(m)], [what type of person], [e_i

[which people] S

voted for e_j

This distinction replicates itself when 'what' and 'how' move in LF as heads of WH phrases.

(c) I wonder who thinks Fred ate what

*(d) I wonder who thinks Fred ate how

A Generalized Binding Theory handles both of these cases by exploiting the fact that what as a referential phrase (R-expression) should be subject to Principle C of the Binding Theory while adjunct phrases like 'why' and 'how' do not have referential properties. (Interested readers should see Aoun (1985) and (Wahl (forthcoming) and chapter 4 for extensive justification of this assumption) and so are not subject to Principle C. This being the case the AGR element in the embedded clause of (a) and (c) counts as an Accessible subject for a 'how' phrase because conindexation of AGR with the subject NP does not induce a Principle C violation. Therefore the embedded clause counts as a binding domain for traces left by the LF movement of 'how' quantifiers in (b) and (d) and since these traces are not bound in this domain the sentences are correctly ruled out. (b) and (d) are grammatical because 'what' being an R-expression does not have a binding domain for reasons discussed above. Thus Pesetsky's examples pose no problem for this approach.

12. An objection to this treatment arises when we consider cases where superiority violations are induced even when the WH element that is crossed over is not (at least superficially) in subject position as in the following examples:

*(a) What did John persuade who to buy?

*(b) What did John expect who to buy?

(a) would not be a problem if we could claim that its structure was as in (c), essentially a binary branching structure as proposed by Kayne (1983).

(c) John persuaded [_S Fred AGR [_S PRO to buy WH]

Evidence that this is the correct structure comes from the consideration of quantifier float, which applies freely when the quantified NP is in subject position, but not when it is in object position as shown by (d).

(d)i. The men all ate a piece of cake.

*ii. John saw the men all for a second.

Notice that the superficial object of a 'persuade' type verb patterns like a subject as shown by (e)

(e) I persuaded the men all to leave.

Chomsky and Larson (personal communication) suggest that the 'all' in this structure may be floating from the PRO subject. If this were possible though, we would expect sentences like (e') and (e'') to be possible, which does not seem to be the case.

(e') *The men tried all to leave.

(e'')* The men wanted all to leave.

(b) is a problem because 'who' is bound in the matrix clause which contains the first accessible subject. In order to handle this case, we adapt a suggestion of Stowell (1982) to exceptional case marking cases. Stowell claims that all clauses have inflection nodes. Thus the structure of (b) is (f).

(f) The men expect[John INFL to buy WH]

Exceptional case marking verbs are exceptional in that they transfer their case features to the lower agreement. Under this assumption, the agreement takes on nominal features in this case and the lower clause thus contains an accessible subject for the WH element.

13. Pesetsky(1985) also argues that subjacency constrains LF movement in cases of non d-linked LF movement. See below for a discussion of this final point.

14. We should mention a few theoretical assumptions that we will use to govern the distribution of non-WH QPs. First, we adopt the extension of the generalised binding framework proposed in Aoun & Hornstein(1985) as a typology of quantifiers. That is, just like full nominal expressions, we claim that quantifiers can be divided into pure A' and A'/C anaphors. Quantifiers like 'every' and 'some' are A anaphors and so we expect that, just like their A-anaphor counterparts, they must be bound in the domain of the first accessible subject. As Aoun and Hornstein note, this captures the fact that these quantifiers, for the most part can only have scope over the sentence that they are contained in. The exception to this generalisation, as in a case like (a), where 'some woman' can have scope over 'every man', falls out from the fact that the embedded clause does not contain a subject given the definitions of 'subject' and 'accessibility' adopted in the previous chapter.

(a) Every man expects some woman to go with him to the party.

This extension of the scope domain to the matrix S parallels the extension of the domain of anaphor binding as shown by

(b).

(b) The men expect each other to go to the party.

As pure A' anaphors, quantifiers like 'every' and 'some' are not subject to Principle C of the binding theory. Since we have crucially assumed that an NP containing a WH QP can act as an accessible subject, we must explain why the quantifier can move from the NP internal determiner position without inducing a Principle A violation for these cases. Notice that our decision to allow this is independent of the cases discussed below. For example, as Higginbotham notes, a pronoun can be interpreted as a bound variable in a structure like (a), indicating that the quantifier must be able to move to a position from which it c-commands the pronoun in this structure. The bound variable interpretation is not possible in a case like (b).

(a) [[_{NP} Every congressman's_i picture]_i makes him_i look important.

(b) *[[_{NP} pictures of every congressman_i]_i make him_i look important.

In order to handle the contrast in these cases, we minimally modify the notion of accessibility, claiming that in order to

be a subject for a category y , an NP must be a subject without " y ". Since all NPs except proper names must have determiners and thus are not NPs or subjects without such determiners, we claim that these NPs are not accessible subjects for QPs in determiner position. For a formalisation of this result, see Weinberg and Hornstein (1986).

15. We are assuming that the binding theory applies in LF only. We have seen above that Generalized Binding must apply in LF after LF movement has applied in order to explain scope assignment. The simplest assumption, therefore, is that the binding theory only applies in LF.

16. Let us assume the definition of Chomsky (1981's) ECP provided by Lasnik and Saito (1984) #23-24.

- α properly governs β iff
- α governs β and
- or
- a. α is a lexical category
 - b. α is coindexed with β .

governs if every maximal projection dominating
also dominates and conversely.

17. Chomsky (1986) pg. 42.

18. Chomsky is forced to say that N governs the subject position if he wants to exclude PRO from this position using the PRO theorem of Chomsky(1981).

19. Actually, this assumption is not necessary because there is no theta relation between the determiner of the NP and the preposition. Chomsky (1986) notes that even though the preposition can govern the specifier position of a category by SPEC-Head agreement, we must not allow it to properly govern this position because otherwise, we could not "...block super-raising"...

Chomsky (1986) pg. 25

20. This assumption is a crucial part of the Barriers account of island violations.

21. The major underlying goal of this work is to show that "...the same categories are barriers in the two cases (government; including antecedent government and Movement (i.e. subjacency ASW))."

22. Huang(1982) also observes that if we allow pied piping in this case, we would incorrectly predict that pied piping from a structure like (a) should also be acceptable as a double question.

*(a) Who said that what was on sale.

These cases would be predicted to be on a par with (b) because both sentences would result in the LF structure (c).

(b) Who said that stories of whose friend were on sale

(c) [_{CP} who_i what_j [_{IP} e_i said that [_{IP} [_{NP} e_j] were on stories of whose friend_j

sale]]]

23. This example is from Cinque(1981) pg.48

24. Sportiche(1981) pg. 226.

25. This is shown by the fact that only the highest NP in this hierarchy can be replaced by the prenominal determiner 'mon', 'ton'...etc.

26. The Romance languages seem to differ with respect to whether genitive NPs can be co-indexed with the determiner position. Thus Chomsky(1986) citing Torrego(1985) claims that even extraction of an agentive PP in Spanish is impossible from the subject position as shown in (a)

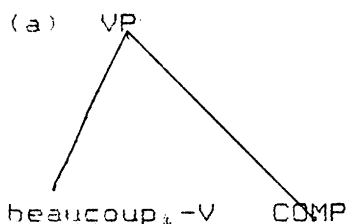
(a) *esta es la autora[_{CP} [de la que_i] [_{IP} [_{NP} varias
this is the author by whom several
traducciones t_i] han ganado premios internacionales.
translations have won international prizes.

Moreover, French differs from Italian in that 'dont' but not 'de qui' can be bound by the determiner. Thus movement of a dont phrase but not a 'de qui' phrase is possible from subject position.

27. All of the data in this section comes from Mouchaweh (1985), who in turn relies on the analysis of Obenauer (1976).

32. Mouchaweh (1985), following Milner (1979) claims that movement is to an adverbial position inside the VP with subsequent cliticisation onto the verb yielding a structure

like (a)



29. Other possibilities are imaginable. We might allow adjunction to a VP by the QP but claim that even though no segment of the internal V^0 dominates the quantifier, some projection of V does dominate and so it is not excluded. While this would allow antecedent government in this case, it would rigidly bar antecedent government into the subject of a small clause position as we will see below.

30. Recall that in the adjoined structure, the NP surrounding the trace of 'combien' does not contain this category and so the verb only governs across one X-max, which is allowed given our definition of lexical government.

31. We observe similar facts about in Italian with respect to 'ne' cliticisation. (See Belletti and Rizzi (1981)). Belletti and Rizzi point out that 'ne' behaves in an analogous manner to 'en' in French. That is, it can be cliticised from both object position, and from the subject position of a small clause. The analysis that we have

proposed for French will also work for 'ne' cliticisation in Italian.

On The Distribution of Parasitic Gaps

7.1 Introduction:

This chapter has three aims: We show how the distribution of parasitic gaps is accounted for within the framework developed in the previous chapters. In particular, we show how the distribution of parasitic gaps is tied to the theory of lexical government proposed in chapter 4. We compare this account to other approaches, and we show the compatibility of this account with the idea that subjacency is a locality condition on the parser; the main idea of chapter three. We will begin by discussing the basic facts that describe the distribution of parasitic gaps.

II. Basic Properties:

The following three properties constrain the distribution of parasitic gaps and should be explained by any theory about these constructions .

1. Parasitic gaps are licensed only by chains that are formed at S-structure. That is, parasitic gaps are not licensed by A-bar chains that are formed at LF.

(1a) Which book_i did you read t_i without reviewing e_i?

(b) *Who read [which book]_i without reviewing e_i

(c) *John read [some books]_i without reviewing e_i.

2. Parasitic gaps cannot be part of a chain containing a c-commanding element in an A-position. This means that chains headed by elements in A-positions do not license parasitic gaps (as shown in (2a)). A-bar headed chains with elements in positions that c-command parasitic gaps also do not license parasitic gaps, as shown in (2b).

(2a) *[The man]_i was rescued t_i without the kidnappers harming e_i.

(b) *Which file t_i got lost before reading e_i.

3. The Subjacency condition constrains the distribution of parasitic gaps. That is, the parasitic gap chain must be subjacent to the real gap chain.²² Thus, as mentioned in Chapter three, parasitic gap chains cannot occur inside syntactic islands.

(3a) Complex NP Violations:

*Which man did you hire t_i without believing [that [the rumor] [that [John would hate e_i]]]

b. Parasitic Gap inside a Relative Clause:

*Which book_i did the publisher accept t_i without meeting
 [NP the author_j [_S who_j [t_j wrote e_i]]]

7.2 Indexation Within the WAHL Framework:

Given the organization of the grammar proposed in Chapter 4, any indexed category that is created by syntactic movement will be fed into the PF component, and will therefore be subject to the ECP. As mentioned in chapter four, this means that we must allow pronominal empty categories to remain unindexed at PF because they occur in ungoverned positions.²⁰ If PRO were indexed at PF these structures should all violate the ECP.

- (4)a. Oh [[PRO to be in England,] now that Spring is here.]
 b. The crowd [was [too angry]] [PRO_{LAT-ED} to hold the
 meeting].
 c. [Pro being a linguist] made Mary happy.

To achieve this result, we followed WAHL(1987) in claiming that lexical (phonetically interpreted items or items with phonetic features) automatically have indices in PF. Syntactic movement also leaves an indexed category in the movement position. If an item is to be interpreted at LF, or if it must be there to

satisfy independently motivated LF conditions, then the category must be indexed at this level. If neither condition is met, then the category can remain unindexed, and uninterpreted at all levels of representation. Note that even though the PRO itself may not be indexed, a trace that it leaves in a position that it moves from will be indexed given the conventions adopted above. We must also reformulate the definition of chain formation slightly, allowing this process to occur between elements whose indices are non distinct instead of forcing linked elements to be coindexed.

The contrasts between movement from a raising and control structure show that this is the correct result.

- (5)a. John [_{VP} asked Fred [_{CP} PRO to be certain [_{IP} t_i to come]]].
 * b. John [_{VP} asked Fred [_{CP} PRO to be probable [_{CP} [_{IP} t_i to come]]]]⁴

7.3 Subjacency and Parasitic Gaps

In the Chapter 5, we discussed the intimate connection between bounding theory and the condition of lexical government at PF. We claimed that the class of permissible landing sites for overt syntactic movement was constrained by the possibility of lexically governing intermediate traces left by syntactic

movement in the PF component. We could expand the class of permissible movement sites for parasitic gaps, if we assumed that the final output of this movement was not subject to the PF ECP. As we saw in Chapter 5, overt movement from adjunct position is marginal and from subject position it is impossible. Both possibilities were ruled out by the interaction of bounding theory and the theory of lexical government.

Since parasitic gaps chains can begin either inside a subject or an adjunct while traces of real movement cannot, we must insure that these structures will not be ruled out by the ECP. We can guarantee this result, if we claim that the "empty operator" at the head of a parasitic gap chain is really a PRO. Chomsky (1982) originally proposed that parasitic gaps were PROs at D-structure. We adopt the spirit of this proposal, claiming that the "parasitic gap" is a PRO at D-structure.¹⁰

The category moves to the head position before PF. Given the WAIII analysis, such PROs are not indexed at s-structure. Since only indexed categories are visible in PF, the "empty operator" structures are automatically exempt from the ECP.

Even though the empty operator itself is unindexed at s-structure, the traces that it leaves as it moves from its d- to S-structure position will be indexed since we assumed that movement always leaves an indexed empty category. This allows us to use the same account that we used to rule out overt movement from syntactic islands for parasitic gap structures. In order for

these cases to satisfy the subjacency condition (by making the empty operator chain subjacent to the real gap chain) the "empty operator" must move to a position where it will be subjacent to a c-commanding element in the real gap chain. This will leave a non-properly governed indexed trace in these cases exactly as non-empty operator movement for the cases that we discussed in chapter 5. The full structures for (3) are provided below and the offending traces are underlined>.

(6)a. * $[_{CP} \text{ Which man}_i \text{ did } [_{IP} \text{ you } [_{VP} t_i \text{ } [_{VP} [_{VP} \text{ hire}] } [_{PP} \text{ without } [_{O}_i \text{ PRO } [_{VP} \text{ believing } [_{NP} [_{NP} \text{ the rumor}] } [_{CP} \text{ } \underline{e}_i \text{ } [_{CP} \text{ } \underline{e}_i \text{ that } [_{IP} \text{ John would } [_{VP} \text{ hate } \underline{e}_i]]]]]]]]]$

b. * $[_{CP} \text{ Which book}_i \text{ did } [_{IP} \text{ the publisher } [_{VP} \text{ accept } t_i]] [_{PP} \text{ without } [_{CP} \text{ } \underline{e}_i \text{ } [_{IP} \text{ PRO } [_{e}_i \text{ } [_{meeting} [_{NP} [_{NP} \text{ the author}_j] } [_{CP} \text{ } \underline{e}_i \text{ } [_{CP} \text{ who}_j \text{ } [_{t}_j \text{ } [_{VP} \text{ wrote } \underline{e}_i]]]]]]]]]]]$

We make crucial use of our ability to adjoin to NP, CP and IP in order to allow parasitic gap constructions to appear inside adjuncts and subject NPs. The category moves to a subjacent position either adjoined to or within the adjunct or subject before PF. Overt movement would result in ungrammaticality because the trace inside the subject or adjunct island would be ruled out by the PF ECP. Since empty operators are unindexed and since only indexed categories are visible in PF, the "empty operator" structures are automatically exempt from the

ECP. Thus (7)a and b, with the structures (c) and (d) satisfy both subjacency and the ECP because the PRO head of the parasitic gap chain is not indexed at PF and so the ECP does not apply to the head of the parasitic gap chain.

(7)a. which picture would admirers of never sell

b. Which movie can you see without discussing

c. [_{CP} which picture would [_{IP} [_{NP} PRO [_{NP} admirers [_{PP} of e_i]]]] never [_{VP} sell t_i]]]]?

d. [_{CP} which movie_i can [_{IP} you [_{VP} see t_i] [_{PP} without [_{CP} PRO [PRO [_{VP} discussing t_i]]]]]]?

Let us first consider the case of (7a) with its structure (7c). In order to be subjacent to the real operator position, the WH element would have to adjoin to the subject NP position. In the case of overt movement, further movement of this trace into SPEC CP would leave an indexed trace which would be ruled out at PF by the condition of lexical proper government. However, in the case of a parasitic gap construction in subject position, the unindexed head of the parasitic gap chain remains in the NP adjoined position. Since this category is unindexed, it is not visible in PF and so escapes the effects of the lexical ECP. This is why overt movement from subject position triggers a CED violation, while a parasitic gap structure is possible from this position.

The same is true of the adjunct position. The head of the parasitic gap chain can remain in the post prepositional position because it does not need to be lexically properly governed. Overt movement from this position is not possible because it would leave an indexed trace that would not survive the lexical ECP.

Notice that in these cases the parasitic gap chain is subjacent to the head of the real gap chain. Therefore, we can claim that subjacency like other grammatical processes including binding, and obligatory control are governed by a c-command condition between licenser and the licensee. We assume, with Chomsky (1986) that the real operator and empty operator chain must form a composed chain at s-structure, and that each element in this composed chain must be subjacent from some element either in the composed or real gap chain. We allow the empty operator (PRO) to leave a set of indexed traces in the positions that it moves from without being indexed itself. As mentioned above, we must reformulate the definition of chain formation slightly, allowing this process to occur between elements whose indices are non distinct instead of forcing linked elements to be coindexed.

This theory can also handle the contrasts first pointed out in Kayne (1983) and given in (8) - (10).

- (8) a. The books you should read before it becomes difficult to talk about.

*b. The books you should read before talking about becomes difficult.

(9) a. A book that he reviewed without believing the first chapter of

*b. A book that he reviewed without believing the first chapter of to be full of lies.

(10) a. A person that close friends of admire

*b. A person who you admire because close friends of became famous

The structure for (9) is given in (11) .

(11)a. [_{NP} the books_i [_{CP} e_i [_{IP} you should [_{VP} [_{VP} read t_i] [_{IP} before [_{IP} PRO [_{IP} it becomes difficult [_{CP} e_i [PRO to [_{VP} talk about e_i]].]]]]]]

b. [_{NP} books_i [_{CP} e_i [_{IP} you should [_{VP} [_{VP} read t_i] [_{IP} before [_{IP} PRO [_{IP} [_{NP} e_i] [_{NP} PRO talking about e_i] becomes difficult.]]]]]]

In order to move to the head of the adjunct, the empty operator must move outside of the NP in the subject position of the adjunct clause. Given our definition of subjacency, the category must first adjoin to the NP. Further movement to IP adjoined position must take place in order for empty operator to be subjacent to the real operator. This movement leaves a non-lexically governed trace, ruling out the structure.

Similar remarks apply to (9) -(10) which also involve movement from a subject of an embedded sentence. The structures are given in (12) and (13)

(12) a. [_{NP} [_{NP} A book_i [_{CP} that_i [_{IP} he_j [_{VP} [_{VP} reviewed t_i] [_{CP} without [_{IP} PRO [_{IP} PRO [_{VP} believing [_{NP} e_i [_{NP} the first chapter of e_i]]]]]]]]]]]

*b. [_{NP} [_{NP} A book_i [_{CP} that_i [_{IP} he_j [_{VP} [_{VP} reviewed t_i] [_{CP} without [_{IP} PRO [_{IP} PRO [_{VP} believing [_{IP} [_{NP} e_i [_{NP} the first chapter of e_i] [to be full of lies]]]]]]]]]]]]].

(13) a. [_{NP} A person_i [_{CP} that_i [_{IP} [_{NP} PRO [_{NP} close friends of e_i] [_{VP} admire t_i]]]]]

*b. [_{CP} [_{NP} A person_i [_{CP} who_i [_{IP} you [_{VP} [_{VP} admire t_i] [_{CP} because [_{IP} [_{CP} e_i [_{NP} close friends of e_i] [became famous]]]]]]]]].

7.4 The Barriers Approach to Parasitic Gaps

Chomsky (1986) proposes to decouple the c-command constraint from the subjacency condition. The main reason for this is the appearance of parasitic gap constructions in subject and adjunct islands; constructions that disallow overt movement. In order to explain this asymmetry, Chomsky (1986) licenses parasitic gaps if they are subjacent to the real gaps left by syntactic movement of a lexical A-bar category.

Notice that the proposal that we have suggested above is incompatible with the Barriers framework because the grammatical structures in (10) – (13) above involve adjunction to NP, CP, and IP. However, as discussed in Chapter 5, Chomsky must rule out adjunction of a category to any of these positions in order to block overt movement out of syntactic islands. Therefore in (10a) – (13a) the parasitic gap will not be subjacent to the real operator. They are separated by two IPs in (10) and by both NP and IP in (11) and (12). The head of the parasitic gap chain is subjacent to the real gap (or the trace of the real gap adjoined to VP). This trace does not c-command the parasitic gap chain.

Therefore, c-command must be removed from the definition of subjacency. The definition of subjacency used by this approach is given in (14).

(14) (b) is subjacent to (a) if there are fewer than $n+1$ barriers for (b) which exclude (a).

We relativize the notion of bounding node to those "categories which are barriers for (b)". Recall from Chapter 5 that a category can only be a barrier for a category that it dominates. Thus subjacency only regulates the number of nodes that contain a category (b) and that separate it from a category (a). There can be an indefinite number of nodes that do not contain (b) but which separate it from its immediate syntactic antecedent (a).

Cases like (15) also make it difficult to interpret subjacency as a constraint on locality. In these cases, the parasitic gap is dominated by a potentially unbounded sequence of L-marked complements. Since L-marking exempts a category from barrierhood, we can again have an unbounded number of categories separating the parasitic gap from the element in the real gap chain that licenses it.

Within the Barriers framework where subjacency is non-symmetrical and does not presuppose c-command, the parasitic gap

trace in the most deeply embedded PP of the subject NP is subjacent to the real gap trace in the adjoined VP position. This is because none of the categories that dominate this NP count as barriers in Chomsky's system because each node L-marks the next. Since all of the categories in the subject NP are complements and so all are L-marked by the nodes that dominate them none of them are barriers.

(15) Which man did articles about proposals for monuments of
enrage?

The structure of (15) as analyzed by the Barriers framework is (16).

(16) [_{CP} which man_i did [_{TP} [_{NP} articles _{PP} about [_{NP} proposals
[_{PP} for [_{NP} monuments [_{PP} of [_{NP} e_i]]]]]] [_{VP} t_i enrage t_j]]]]?

In Chapter 3, we explained the fact that subjacency applied only to s-structure gaps by interpreting this condition as a locality constraint imposed by efficient parsing. In that Chapter we showed that locality was crucial in order to allow a deterministic and unambiguous parse of sentences with ambiguous complement structures. The ambiguity in these cases came from the fact that since traces are not phonetically realized, the parser has no overt local cue to tell it how to expand a verb with multiple possible complement structures.

We will repeat one of the ambiguous structures in (17)

- (17)a. Which cake_i did you bake t_i instead of eating e_i?
 b. Did you bake a cake today instead of eating?

Through use of the theory of LR(k) parsing, we were able to show that efficient parsing could be guaranteed only if we could establish a bound on the Left context over which a parser would have to search for disambiguating information. LF movement and other grammatical phenomena, by contrast leave overt syntactic cues that instruct the parser in how to expand a syntactic structure. Notice that if we adopt the Barriers interpretation of the subjacency condition, we can no longer maintain this explanation because this definition does not bound the search space for disambiguating context. A bounded number of categories containing a potential parasitic gap position plus an unbounded number of non-containing categories may potentially contain the parasitic gap or empty operator's licenser. Given that adopting this definition of subjacency forces us to lose an explanation for why subjacency only applies s-structure processes, we will continue to pursue the definition proposed in Chapter 5 of this thesis. In the next section, we provide additional support for the idea that the head of a parasitic gap chain is an unindexed PRO.

7.5 Anti C-Command and Parasitic Gap Licensing

In this section, we will discuss the anti c-command requirement and derive it from the notion of "local A-bar binder" presupposed by GEBI.

Consider a case like (18)

(18) Which man_i does he think [_{CP} t_i [John likes t_i]

Notice that strong crossover effects apply in this case and the variable cannot be coindexed with the pronoun in the matrix clause. Within the GEBI framework, we do not delete intermediate traces at LF. Therefore we must state Principle C in a way that does not allow the semantically null intermediate trace to protect an *c*-expression from the effects of this condition. In order to do this restate Principle C as in (19).

(19) A category must be free in the domain of a semantic local A'-binder.

We interpret a semantic A'-binder to be an element that plays a role in the semantic interpretation of the variable position. A normal WH element binds a variable and the restrictor determines the variable's range. The empty operator in a predication structure identifies the empty position as a variable position and binds the position creating a predicate that can then be predicated of another position in the structure. The trace in

(18) does neither of these things. It does not semantically bind the variable or restrict its range. Similarly, in parasitic gap constructions, the empty operator does not bind the variable, nor does it restrict its range. Therefore it should be treated like an intermediate trace in (18), i.e. a non semantic binder that does not protect the variable from Principle C violations.

With this in mind, consider a case like (20)

(20) Who did you meet t_i without greeting e_i ?

In Chapter 3, we showed that empty parasitic gaps are not c-commanded by the elements in A-positions of the real gap chain whose operator they are linked to. Cases like (21) show that empty operators must not be part of a composed chain that has c-commanding elements in A-positions.

(21a) * $[\text{The man}]_i$ was rescued t_i without the kidnappers harming e_i

(21b) *Which file t_i got lost before reading e_i

Therefore, the structure of a typical parasitic gap construction fed to the LF component will be (22) ¹⁹

(22) [_{CP} Which file_i [_{IP} t_i [_{OP} [_{OP} got lost] [_{without} [_{CP} [_{IP} PRO_i
 [_{OP} reviewing e_i].

In this structure, the PRO-empty operator is not a semantic A-bar binder and so it cannot protect the parasitic gap from being interpreted as directly bound to the real gap. The real gap however is in an A position that c-commands the parasitic gap. Since the parasitic gap is a variable and is bound to an A-position in this structure, we induce a Principle C violation and the structure is ruled out.

Similar remarks apply to cases where the head of the chain is in an A-position that c-commands the parasitic gap. Here again, since the PRO head of the parasitic gap chain is a non semantic A-bar binder, the parasitic gap is A-bound to the head of the real gap chain.

Browning (1987) points to a set of constructions that illustrate the effects of a semantic A-bar binder. She notices that the anti c-command properties do not apply to variables within so-called Predication structures. We will repeat some of her examples to illustrate this point.

(23) a. What_i did you [_{OP} give the kids t_i [_{OP}-PRO_i [_{CP} PRO_i to
 play with e_i]]]

b. Who_i e_i [_{CP} is easy [_{PRO-OP}_i [_{PRO}_j to please e_j]]]]¹¹

Predication operators bind open positions within their predicate phrases. We can even treat these as lambda operators in the sense of Williams (1977). Since these operators have a semantic function, they count as local A-bar binders, and so protect the variable in the parasitic gap chain from being locally A-bound to a c-commanding category in an A-position.

We follow Aoun and Clark (1983) in claiming that these structures receive a predication index that is different from the NP that the clause is predicated of. This process applies to the formerly unindexed empty operator that was part of the subjacency chain in 23a. and b. Since this predication index applies after subjacency is checked, we form a composed chain between the real gap and empty operator chain for the purposes of subjacency¹². The Predication rule applies at LF, after the binding theory is checked thus allowing cointerpretation between the c-commanding real operator and the parasitic gap in the Predication structure. The main point is that distinct predication operators will be present at LF. Since the LF structure (23) is (24), we predict these cases obey GEBI restrictions at LF. Since the parasitic gap is bound to an semantic empty operator, Principle C is not violated in these structures.

(24) [_{CP} what toy_i did [_{you} give t_i the kids [_{PRO-OP}_i [_{PRO}_m to play with e_m]]]].

Reindexing applies under predication and so the final interpreted structure is (25).

- (25) [_{CP} what toy_i did [_{you} give the kids t_i [_{CP} PRO_i [_{PRO_m} to play with e_i]]]]

This argument will apply to predication operations like purposives and tough movement constructions in general, as the reader can verify for himself.

Assuming that the empty operator is a PRO rather than some non-NP type category immediately allows us to explain why parasitic gaps are not licensed by PPs as first pointed out in Chomsky (1982)¹⁰ and discussed extensively in Cinque (1981). If PROs are NPs, then we predict that a sentence like (26a) or (26b) will have no d-structure source.

- (26) a. * I saw a friend_i [_{PP} about whom]_i John_j talked without PRO_j caring [_{PP} e_i]

7.6 S-Structure Licensing

As mentioned at the beginning of this paper, parasitic gaps must be licensed at s-structure.

The theory developed so far accounts for this fact. That is, LF chains that end in real gaps in A-positions that c-command the parasitic gap will induce Principle C violations, given our assumption that the PRO-empty operator in a parasitic gap structure is a non semantic operator. Thus (27a) and (27b) are ruled out.

- (27) a. Every book_i x_i fell off the table without reading e_i.
 b. Who asked which book_i x_i fell off the table without reading e_i.

Recall that we have assumed that subjacency incorporates the c-command condition. In Chapter two of this thesis, we argued that elements in a non c-commanding relation are placed in separate parsing substacks and that material must be in the same substack as a category in order to influence parsing decisions about the creation or placement of that category. This means that in order to create an empty operator to begin a parasitic gap chain, we must search for a c-commanding antecedent. Thus cases like (28) will be ruled out because, since the *wh* in-situ does not c-command the empty operator position until LF, the operator would not be created at S-structure. There would be no basis to create the entire parasitic gap chain on the assumption that empty categories can only be created if there is a c-commanding potential antecedent in a subjacent position.

- (28) a. *Who read [which book]_i without reviewing

- b. *John read [some books]_x without reviewing

7.7. A Surprising Assymetry

Consider the sentences in (29)

- (29) a. Which book did John review without believing could be a
bestseller?
b. *Which book did John review before could become a
bestseller?

Our theory predicts the assymetry between these cases in a straightforward way. As usual, we assume that both parasitic gap chains are headed by PROs. The PRO in (b) cannot be indexed at s-structure because, since prepositions are not proper governors, it would be ruled out by the PF ECP. This means though, the subject trace of the adjunct will be ruled out by the PF ECP given the assumption that prepositions are not proper governors and that the PRO in SPEC CP, being unindexed at PF cannot transfer its features to the head of CP, allowing this head to act as a proper governor. The PF structure is (30).

- (30) Which book_x did John review t_x [_{PF} before [_{CP} Pro [t_x could become a bestseller]]].

By contrast, the trace in the embedded subject position is properly governed by the verb 'believe'. Therefore, even though the PRO at the head of the adjunct must be unindexed at PF, the structure will still satisfy the PF-ECP as (31) shows.

(31) Which book_i did John_i review t_i [_{CP} before [_{CP} [_{IP} PRO believing [_{IP} [_{IP} e_i could be a bestseller]]]].

7.8 Other Approaches:

We will compare our account with that of Aoun and Clark (1983) and Browning (1987).¹⁴

Our account borrows Browning's idea that the parasitic gap operator, as opposed to the predication operator does not protect the variable in the parasitic gap structure from being A-bound. Browning derives this restriction by claiming that empty operators heading parasitic gap chains are simply invisible at LF because they cannot be indexed at this level. She claims that indexed empty operators are uniformly interpreted as signs of predication at this level and thus, if we indexed the head of the parasitic gap chain, the structure would receive the wrong interpretation. While this account seems to derive a difference that we simply state in terms of the notion of "semantic A-bar binder", on further inspection, this derivation seems illusory. First of all, lexical operators seem subject to a variety of

semantic interpretations. A 'wh' element in a question is an operator that binds a variable and closes an open sentence. The 'wh' element in a predication structure is a lambda abstractor that creates an open sentence that can be predicated of another position. Since overt Wh elements can receive a variety of interpretations at LF, it seems ad hoc to restrict the interpretation of empty operators at this level.

Our account differs from both that of Aoun and Clark's and Browning's in that we derive the full range of anti c-command effects by incorporating the c-command restriction directly into the subjacency constraint while Browning divorces these two restrictions and claims that c-command is not relevant to bounding theory. It is only relevant the notion of an antecedent or binder in that an antecedent must c-command a category that it binds. Aoun and Clark claim that generalized binding enforces the locality restriction between real and parasitic gap chains. We believe that our approach, which treats locality restrictions as a function of subjacency governed by c-command restrictions is a more adequate approach on both conceptual and empirical grounds.

There are empirical problems with the approach that Aoun and Clark adopt, pointed out by Browning (1987) and noted above. Besides these empirical problems, Aoun and Clark disallow the licensing of parasitic gaps by LF movement by claiming that the empty operator is an A-anaphor and that the Binding Theory applies

at S-structure, as well as LF. Since a non c-commanding WH in situ in a case like (28) will not be in the appropriate position to bind the operator until LF, these structures will be ruled out at S-structure.

This assumption is problematic because for all other cases, generalized binding applies at LF. The child therefore has no evidence for assuming that the condition also applies at s-structure except for cases like (28). Since it is not reasonable to assume that such sentences are part of the PLD, we assume that the child could mistakenly assume such cases to be grammatical because they would obey generalized binding at LF. Since we need to assume independently that subjacency applies at s-structure and can motivate why this is so, it seems better simply let this condition apply at s-structure and let the Generalized Binding theory apply uniquely at LF.

Browning (1987) employs the Barriers definition of subjacency, so she also cannot use this condition to rule out cases like (21) and (22) because the empty operator is in fact subjacent to the quantifier or wh in situ in these cases. She rules these cases out by adopting a strong binding condition which she also must claim applies at s-structure in the case of parasitic gaps. For parasitic gaps, strong binding involves linking the parasitic gap chain to a lexical operator that can bind the chain. This stipulation is ad hoc because again, parasitic

gaps are the only case that force strong binding to apply at s-structure. For all other cases, we must only assume linkage to a lexical operator for identification purposes in some interpretive component. In addition, strong binding must be interpreted disjunctively because the predication structures discussed above are not linked to any lexical category until a post LF level. The index assigned to the predication structure is semantically meaningless until this level. Therefore, it is unclear why the predication operator should be a better strong binder than an unindexed operator. We are left with a disjunctive and peculiar definition of "strong binding".

Moreover, this condition can only apply to parasitic gaps because some variables are not strongly bound until LF.

This is the case in (32), where the quantifier that strongly binds the pronoun, turning it into a bound variable does not even c-command the pronoun until LF.

(32) a. Mary [_{CP} canvassed every senator_i] before he_i voted on Contra aid.

b. Every x_1 : Mary canvased x before x voted on Contra aid.

It is also not clear that the c-command restriction can be simply seen as a fact about binding, as Browning supposes, because

this restriction seems to apply to the gapping examples cited in chapter Three. In Chapter Three, we discussed the contrast in cases like (33) below. The information about inflection has to reach a position where it c-commands the gap in the second conjunct, but here we would not be tempted to say that this information binds this position.

(33) I believe Fred ran after Mary and Bill after Sue.

* b. I believe that Fred ran after Mary and that Bill after Sue.

Finally, if we assume that strong binding applies to parasitic gaps at s-structure, but to WH traces at LF, then we predict that there is a possible language where the WH element can move to a subjacent non c-commanding position at s-structure and then to move at LF to a position where it strongly binds the variable. We know of no such language. (34) is such a structure. In this case, we move the WH element into the non-theta marked sentential subject position first. Then we move it at LF into the CP.

(34) who said that for who to be believed that Bill left amused.

7.9 Conclusion

We have tried to show that we can use the assumptions in WAHL to derive the fact that the empty operator in a parasitic gap

construction bears no index at LF. We used this fact and the WAHL theory of ECP proper government to derive the fact that parasitic gap constructions cannot be headed by A-chains, nor by c-commanding WH in situ quantifier constructions. The WAHL framework also predicts ungrammaticality of parasitic gaps from subject position. The theory predicted that parasitic gaps could occur inside CED structures. The theory also predicted the unacceptability of parasitic gaps in traditional syntactic islands. Moreover, we argued that the definition was conceptually more appealing than the non symmetric definition of Chomsky (1986) and Browning (1987) because it allowed us to explain why subjacency was an s-structure phenomenon; in fact the only s-structure condition in the grammar. We also motivated a parsing theoretic c-command condition which we use to explain why LF A' chains couldn't license parasitic gaps and showed that it compared favorably with Aoun and Clark or Browning's explanation for this fact. We take this analysis as independent support for the WAHL framework, and particularly for the division of the ECP into both a PF and LF condition. We also hope that this is case demonstrates that conditions imposed by the parsing as well as the learning mechanism can explain a subtle range of syntactic properties.

Footnotes

1. Following standard conventions, we will indicate the "real" gap of WH-movement by "t" and the "parasitic" gap by "e".

2. We will argue below that the subjacency requirement is satisfied only if the head of the parasitic gap chain is subjacent to the head of the real gap chain. We will compare this treatment with approaches claiming that the subjacency relation holds between the empty operator and the real gap.

3. If we assume Chomsky (1981)'s derivation of the PRO theorem, we are led to conclude that PRO can only occur in ungoverned positions.

4. Notice that we must allow PRO to be indexed at LF or else cases like (a) will violate the Binding Theory. This was pointed out to me by Juan Uriagareka.

(a) [[PRO_{LF} teaching oneself mathematics] can be difficult.

We will see below that while this type of indexing must be allowed at LF, we must block arbitrary indexing or reindexing at this level. See below for further discussion.

5. The major difference between this proposal and Chomsky's is that Chomsky claimed that the pro-parasitic gap was directly bound to the WH-operator and thus became a variable at later levels of representation. This procedure presupposes that empty categories are functionally defined; a view that has been shown to be incorrect by Brody (1984). In addition, as Chomsky (1986) notes, since the parasitic gap is not directly subjacent to the real gap or its binder, we cannot claim that parasitic gap constructions are governed by subjacency and thus lose the account of why parasitic gap constructions cannot appear inside islands. Note that we can claim that a PRO can begin in the post verbal position as long as it does not remain in this position at s-structure, or PF. This is because PRO is barred from this position by the PRO theorem of the Binding theory under Chomsky's treatment of the distribution of PRO, or by case theory, assuming the Bouchard (1984), Hornstein-Lightfoot (1987), Manzini (1983), or Sportiche (1983) theory of PRO.

6. Recall, that adjoining the trace of the empty operator to Cp breaks up the mutual government needed for predication. Thus, although the adjoined trace is lexically governed, the relative clause cannot be predicated of its head.

Bibliography

Aho, A.V. and JD Ullman (1972) The Theory of Parsing, Translation and Compiling. Prentice Hall.

Andrews, A. (1982) "A Note on the Constituent Structure of Adverbials and Auxilliarities". Linguistic Inquiry 13.2

Aoun, Y. (1981) On the Formal Nature of Anaphoric Relations unpublished doctoral dissertation MIT.

Aoun, Y. (1985) A Grammar of Anaphora. MIT Press.

Aoun, Y. and R. Clark (1983)

Aoun, Y. and N. Hornstein (1985) "Quantifier Types". Linguistic Inquiry. 16.4.

Aoun, Y., N. Hornstein, D. Lightfoot, and A. Weinberg (forthcoming) "Two notions of Locality. Linguistic Inquiry 18.

Aoun, Y. and D. Lightfoot (1984) "Government and Contraction." Linguistic Inquiry 15.

Aoun, Y. and D. Sportiche (1983) "On the formal theory of government". The Linguistic Review 3.

Aoun, Y., N. Hornstein and D. Sportiche (1981) "On some aspects of wide scope quantification." Journal of Linguistic Research 1.3.

Baker, C. L. (1970) "Notes on the description of English questions: the role of an abstract question morpheme. Foundations of Language 6.

Barton, E., R. Berwick and E. Ristad (1987) Computational Complexity and Natural Language. MIT Press.

Belletti, A. and L. Rizzi (1981) "The syntax of 'ne': some theoretical implications". The Linguistic Review 1.

Berwick, R. (1985) The Acquisition of Syntactic Knowledge. MIT Press.

Berwick and Weinberg (1984) The Grammatical Basis of Linguistic Performance. MIT Press.

Berwick and Weinberg (1985) "Deterministic parsing and Linguistic Explanation". Language and Cognitive Processes 1.2

7. At the moment, we are only discussing the s-structure licensing of the parasitic gap chain and how this chain meets PF conditions. We will discuss the binding of the parasitic gap below.

8. We assume that the 'of' is inserted in the subject NP of 22c. after the ECP is checked. We also assume that VP adjoined traces are properly governed by preposing the verb to INFL position, as suggested by Chomsky (1986). With WAHL, we assume that local head movement doesn't leave an indexed empty category.

9. Recall that we must assume that both lexical projections of major categories can reanalyse prepositional phrases that they govern and thus properly govern the objects of these prepositional phrases.

10. Irrelevant details are omitted.

11. These examples are from Browning (1987) pg. 2, #13.

12. The structure of this chain is given below (irrelevant details omitted).

(i) [_{NP} what_i did [_{IP} you [_{VP} t_i [give the kids [_{PRO} [_{PRO}_i to
[play with e_i]]]]]]]

13. Chomsky (1982) attributes this observation to Pesetsky.

14. We have nothing to add to Browning's critique of Stowell (1985), who also deals with parasitic gap constructions in a way that is inconsistent with this approach. We refer the reader to Browning's work on this topic.

Bouchard, D. (1984) On the Content of Empty Categories. Foris

Bresnan (1972) The Theory of Complementation in English Syntax.

MIT PhD dissertation.

Brody, M. (1984) "On contextual definitions and the role of chains". Linguistic Inquiry 15.

Browning, M. (1987) "Null Operator Constructions". MIT ms.

Chomsky, N. (1973) "Conditions on Transformations". A Festschrift for Morris Halle. Holt Rinehart.

Chomsky, N. (1975) Reflections on Language. Pantheon.

Chomsky, N. (1981) Lectures on Government and Binding. Foris.

Chomsky, N. (1982) Some Concepts and Consequences of the Theory of Government and Binding. MIT Press.

Chomsky, N. (1986) Barriers. MIT Press.

Cinque (1981) "On Extraction from NP in Italian". Journal of Italian Linguistics. 5_

Contreras, H. (1985) "Note on Parasitic Gaps". Linguistic Inquiry

Crain and Fodor (1984) "How Grammars Can Help Parsers. Natural Language Parsing: Psychological, Computational and Theoretical Perspectives. D.R. Dowty, L. Karttunen, and A. Zwicky, eds. Cambridge University Press

Engdahl, E. (1981) "Multiple Gaps in English and Swedish". in T. Freitheim and L. Hellan, eds. Proceedings of the Sixth Scandinavian Conference of Linguistics. Tapir.

Evans, G. (1980) "Pronouns". Linguistic Inquiry 11.

Fodor, J. (1985) "Deterministic parsing and Subjacency ". Language and Cognitive Processes. 1.1

Frazier, L. and J.D. Fodor (1978) "The Sausage Machine". Cognition 6.

Frazier, L. and K. Rayner (1982) "Making and Correcting Errors During Sentence Comprehension: Eye Movements and the

Analysis of Structurally Ambiguous Sentences. Cognitive Psychology 14.

Gorrell, P. (1987) Studies of Human Syntactic Processing - Ranked Parallel versus Serial Models. University of Conn. PhD thesis.

Haik, I. (1985) The Syntax of Operators. unpublished MIT PhD diss.

Heim I. (1982) The Semantics of Definite and Indefinite Noun Phrases. University of Massachusetts GLSA

Heny, F. (19xx) Binding and Filtering

Higginbotham, J. (1980) "Pronouns and Bound Variables".
Linguistic Inquiry 11.4

Higginbotham, J. (1985) "On semantics" Linguistic Inquiry 16.

Higginbotham and May (1981) "Questions, quantifiers, and crossing". The Linguistic Review 1.

Hornstein, N. (1984) Logic as Grammar. MIT Press.

Hornstein, N. (submitted) "Restructuring in the T model".

Hornstein, N. and D. Lightfoot (1987) "Predication and pro".
Language 63.

Hornstein, N. and D. Lightfoot (forthcoming) "On lexical
government". R. Freidin, ed. Princeton Conference Proceedings.

Hornstein, N. and A. Weinberg (1981) "Case theory and preposition
stranding". Linguistic Inquiry 12.

Hornstein, N. and A. Weinberg (1987) "Superiority and generalized
binding". Proceedings of NELS XVII.

Huang, J. (1982) Logical Relations in Chinese and the Theory of
Grammar. MIY PHd diss.

Jaeggli, O. (1980) "Remarks on "to" contraction". Linguistic
Inquiry 11.

Johnson, K. (1987) "Against the notion SUBJECT". Linguistic
Inquiry 18

Kayne, R. (1981) "ECP Extensions". Linguistic Inquiry 12.

Kayne, R. (1983) Connectedness and Binary Branching. Foris

- Kenstowicz, M. (1985) "The phonology and syntax of wh-expressions in Tangale". Studies in the Linguistic Sciences 15. U. of Illinois.
- Kimball, J. (1973) "Seven principles of surface structure parsing in natural language". Cognition 2.
- Knuth, D. (1965) "On the translation of Languages from left to right". Information and Control 8.
- Koopman, H. The Syntax of Verbs: From Verb Movement Rules in the Kru Languages to Universal Grammar. Foris.
- Koopman, H. and D. Sportiche (1982) "Variables and the bijection principle". The Linguistic Review 2.
- Koopman, H. and D. Sportiche (1986) "A note on long extraction in Vata and the ECP". Natural Language and Linguistic Theory 4.
- Lasnik, H. and J. Kupin (1977) "A restrictive theory of transformational grammar". Theoretical Linguistics. 4.3
- Lasnik, H. and M. Saito (1984) "On the Nature of proper government". Linguistic Inquiry 15.2.

Lewis, H. and C. Papadimitriou (1981) Elements of the Theory of Computation. Prentice Hall.

Lightfoot, D. (1986) "A brief response". Linguistic Inquiry 17.

Manzini, M.R. (1983) "On control and control theory". Linguistic Inquiry 14.

Marcus (1980) A Theory of Syntactic Recognition for Natural Language. MIT Press.

Marcus, M., D. Hindle and M. Fleck (1983) "Talking about talking about trees. Proceedings of the 21st Meeting of the Association for Computational Linguistics.

May, R. (1985) Logical Form. MIT Press.

Miller, G. (1951) Language and Communication. McGraw -Hill.

Milner, J. C. (1978) De la Syntaxe a l'Interpretation: Quantites, Insultes, Exclamations. Seuil.

Mouchaweh, L. (1985) "De la quantification a distance et des nominalisations en francais. Modeles Linguistiques VII.

Obenauer, H. (1976) Etudes de Syntaxe Interrogative du Français. Niemeyer.

Pesetsky, D. (1981) "Complementizer -trace phenomena and the Nominative Island Condition. The Linguistic Review 1.

Pesetsky, D. (1985) "Wh in situ". unpublished U MASS ms.

Reinhart, T. (1983) Anaphora & Semantic Interpretation. Croom-Helm.

Riemsdijk, H. van (1978) A Case Study in Syntactic Markedness: the Binding Nature of Prepositional Phrases. Foris.

Rizzi, L. (1982) Issues in Italian Syntax. Foris

Sobin, N. (1987) "The variable status of Comp- trace phenomena". Natural Language and Linguistic Theory 5.

Sportiche, D. (1981) "Bounding Nodes in French" The Linguistic Review 1.

Sportiche, D. (1983) Structural Invariance and Symmetry in Syntax. MIT PHd diss.

Stillings, J.T. (1975) "The formulation of gapping in English as evidence for variable types in syntactic transformations."

Linguistic Analysis 1.

Stowell, T. (1981) Origins of Phrase Structure. MIT PhD diss.

Stowell, T. (1985) "Licensing conditions on null operators".
unpublished UCLA ms.

Tanenhaus, M., G. Carlson and M. Seidenberg (1985) "Do listeners compute linguistic representations? in D. Dowty et al, Natural Language Parsing: Psychological, Computational and Theoretical Perspectives. Cambridge University Press

Torrego, E. (1985) "On empty categories in nominals" unpublished
UMass, Boston mss.

WAHL (forthcoming) see Aoun, Y., N. Hornstein, D. Lightfoot and A.
Weinberg

Wanner and Maratsos (1978) "An ATN approach to comprehension"
Linguistic Theory and Psychological Reality. MIT Press.

Weinberg, A. and N. Hornstein (1986) "The necessity of LF".
unpublished UMD ms.

Williams, E. (1977) "Discourse and logical form". Linguistic Inquiry 8.

Williams, E. (1978) "Across-the board rule application." Linguistic Inquiry 9.

Williams, E. (1980) "Predication". Linguistic Inquiry 11.

Williams, E. (1983) "Against small clauses". Linguistic Inquiry 14.

Williams, E. (1986) "A Reassignment of the Functions of LF". Linguistic Inquiry 17.2.

Wright and Garrett(1984) "Lexical decision in sentences: effects of syntactic structure". Memory and Cognition. 12

Zubizarretta, M.L. (1982) On the Relationship of the Lexicon to Syntax. MIT PHd diss.

BIOGRAPHICAL NOTE

Amy Weinberg was born in Boston, Massachusetts. She attended McGill University from 1973-1976 where she received a joint first class honours degree in Linguistics and Philosophy. She has been on the faculty of the University of Maryland since 1984, first in the linguistics department and then on a joint appointment in linguistics and in the university's Institute for Advanced Computer Studies. She has also taught at the Salzburg Summer Institute of Linguistics, and at the Universite de Quebec a Montreal. She serves on the editorial board of Cognition. Her publications include:

(1984) The Grammatical Basis of Linguistic Performance. MIT Press. (with R. Berwick)

(1981) "Case Theory and Preposition Stranding". Linguistic Inquiry 12.1. pp. 55-92. (with N. Hornstein)

(1987a) "Modularity and the Syntactic Processor". Modularity in Knowledge Representation and Natural Language Understanding. Jay Garfield, ed. MIT Press.

(1987b) "Psychological Theory and Linguistic Explanation." Attention and Performance 1986. Max Coltheart, ed. Lawrence Erlbaum.

(1987c) "Two Notions of Locality." Linguistic Inquiry 18.4 pp. 537-578. (with J. Aoun, N. Hornstein, and D. Lightfoot).

and numerous other articles.