THE ACQUISITION OF THE ENGLISH AUXILIARY SYSTEM AND ITS RELATION TO LINGUISTIC THEORY

By

HENRY DAVIS

B.A., The University of Cambridge, 1980

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Department of Linguistics)

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

October 1987

© Henry Davis, 1987

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of LINGUISTICS

The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date 15/10/1987

DE-6(3/81)

ABSTRACT

Thesis supervisor: David Ingram

This thesis explores the connection between linguistic theory, as embodied in a version of the Government - Binding (GB) model of syntax, and the parameter-setting theory of language acquisition.

In Chapter 2, it is argued that by incorporating the criterion of epistemological priority, syntactic theory can move closer towards becoming a plausible model of language acquisition. A version of GB theory is developed which adopts this criterion, leading to several modifications, including the derivation of X-bar theory from more "primitive" grammatical sub-components, and a revision of the Projection Principle. This model is converted into a procedure for phrase-structure acquisition, employing sets of Canonical Government Configurations and Percolation Principles to map Case- and θ-relations onto phrase-structure trees. The chapter ends with a discussion of the "missing-subject" stage in the acquisition of English.

Chapter 3 concerns auxiliaries. It is argued that parametric variation in auxiliary systems can be reduced to levels of association between INFL and V. The question of irregularity is dealt with through the Designation Convention of Emonds (1985), which makes a distinction between open- and closed- class grammatical elements, and a Parallel Distributed Processing

model of learning. The last part of the chapter investigates the learning of the English auxiliary system, and in particular the errors known as "auxiliary overmarking".

Chapter 4 investigates the syntax of Subject Auxiliary Inversion (SAI)-type rules. An account of inversion is developed based on the theory of predication, in which inversion-inducing elements are treated as "A'-type" subjects which must be linked to AGR in order to satisfy conditions on Predicate-licensing. A parametrization is developed based on the cross-linguistic examination of SAI-type rules.

Chapter 5 concerns the acquisition of SAI. It is argued that there are no invariant "stages" in the development of inversion; rather, a proportion of children misanalyze (WH + contracted auxiliary) sequences as (WH + AGR-clitic) sequences and formulate grammars in which SAI is unnecessary. A "two-tiered" theory of syntactic acquisition is proposed to account for the observed developmental patterns.

TABLE OF CONTENTS

Abstract ----------ii

Chapter 1: Introduction

1.0 Preliminaries

1.00 Introductory Remarks----------1

1.01 Learnability and Language Acquisition----------3

1.02 Universal Grammar, Principles, and Parameters---------8

1.03 The Role of Evidence in Language Acquisition---------11

1.04 Acquisition Mechanisms----------16

1.05 Summary----------19

1.1 Parameter Setting

1.10 Introduction----------19

1.11 Ordering the Parameters: the Linear-parametric Model-20

1.12 Parameters and Learnability----------21

1.13 Parameters and Cross-Linguistic Markedness----------28

1.14 Parameters and Real-Time Language Acquisition--------36

1.2 Variability

1.20 Introduction----------38

1.21 Inter-Subject Variability----------38

1.22 Intra-Subject Variability----------42

1.23 Type Variability----------43

1.24 Token Variability----------50

1.25 Conclusion----------53

1.3 Performance Constraints and First-Language Acquisition

1.30 Introduction----------53

1.31 Perception Factors----------56
1.32 Production Factors----------59
1.33 Interpretive Complexity Factors----------65
1.34 Interference Factors----------67
1.35 Conclusion----------68
1.4 On Methodology
1.40 Introduction----------69
1.41 Naturalistic Studies----------69
1.42 "Enhanced Production" Techniques----------71
1.43 Imitation Techniques----------72
1.44 Comprehension Techniques----------74
1.45 Judgement Techniques----------75
1.46 Longtitudinal Versus Cross-Sectional Studies----------76
1.47 Task-Specificity and the Cross-Methodological Approach----------78
1.48 Conclusion----------80
1.5 Outline of the Thesis
1.50 Introduction----------81
1.51 Chapter 2----------81
1.52 Chapter 3----------83
1.53 Chapter 4----------84
1.54 Chapter 5----------85
1.55 Conclusion----------85

Chapter 2: GB Theory as a Theory of Language Acquisition

2.0 Introduction----------87
2.1 Grammatical Subcomponents

2.10 Architecture of the GB Model----------88

2.11 The Projection Principle----------89

2.2 Government

2.20 Overview----------91

2.21 Internal Government----------92

2.22 External Government----------97

2.3 θ-Theory

2.30 The θ-Criterion----------101

2.31 On PRO----------106

2.32 θ-Government----------113

2.33 Predication----------117

2.34 Summary----------121

2.4 Case Theory

2.40 Argument Visibility----------122

2.41 Internal Case-Assignment----------123

2.42 External Case-Assignment----------126

2.43 The Ergative Case-Marking Parameter----------129

2.5 X-bar Theory and Categorial Features

2.50 X-bar Theory and Restrictiveness----------134

2.51 X-bar Theory and Epistemological Priority----------137

2.52 Principles of Tree Construction----------140

2.53 Categorial Features and Categorial Association------141

2.54 Syntactic Licensing Conditions----------145

2.55 Adjectives and Feature Percolation----------147

2.56 Principles of Categorial Feature Percolation--------150

2.57 Deriving Percolation Principles----------162

2.58 Categorial Feature Percolation and the MOLLUSC

Conditions----------164
2.6 Binding Theory----------169
2.7 Bounding Theory----------173
2.8 The ECP----------180
2.9 Learnability, Epistemological Priority and the GREPP
2.90 Introduction----------189
2.91 Government and Epistemological Priority----------190
2.92 Government and Language Acquisition----------209
2.93 Pinker's Model of the Acquisition of Phrase-Structure----------212
2.94 O'Grady's Model of the Acquisition of Phrase-Structure----------222
2.95 Subjectless Sentences and the Pro-Drop Parameter----230
2.96 Conclusion and Summary of Chapter 2----------240

Chapter 3: On Auxiliaries

3.0 Introduction----------243
3.1 Auxiliary Restrictions
3.10 Ordering Restrictions----------246
3.11 Morphological Restrictions----------247
3.12 Subject Auxiliary Inversion----------248
3.13 Sentence Negation----------249
3.14 Floating Quantifiers----------250
3.15 Adverbs----------250
3.16 VP-rules----------251
3.17 Imperatives----------254
3.18 Complementation----------255

3.19 Other Constituency Tests----------262

3.2 On the Category Status of Auxiliary Elements

3.20 Modals----------264

3.21 Auxiliary Do----------267

3.22 Auxiliary Have and Be----------269

3.3 On the Association Between [I] and [V]

3.30 Visibility----------275

3.31 G-Visibility, Full Interpretation, and [I]/[V] Association----------276

3.32 Levels of [I]/[V] Association----------280

3.33 PF Association----------281

3.34 Syntactic Association----------283

3.35 Lexical Association----------284

3.36 Do and Levels of Lexical Insertion----------286

3.4 Irregularity, Lexical Insertion, and the Designation Convention----------288

3.5 Auxiliary Restrictions Revisited

3.50 Ordering Restrictions----------295

3.51 Morphological Restrictions----------298

3.52 Subject Auxiliary Inversion----------299

3.53 Sentence Negation----------299

3.54 Floating Quantifiers----------300

3.55 Adverbs----------300

3.56 VP-rules----------301

3.57 Imperatives----------305

3.58 Complementation----------305

3.6 The Auxiliary System Cross-linguistically

3.60 Introduction----------307
3.61 INFL in German----------308
3.62 INFL in Scandinavian----------323
3.63 INFL in Romance Languages----------326
3.64 INFL in French----------335
3.7 Parametrizing the Model
3.70 Introduction----------337
3.71 INFL as a Universal Category----------338
3.72 Parameters of [I]/[V] Association----------340
3.73 [I]/[V] Association and Learnability----------347
3.8 The Acquisition of the English Auxiliary System
3.80 Introduction----------351
3.81 Overview of Auxiliary Acquisition in English--------352
3.82 Auxiliary Overmarking----------359
3.83 Conclusion----------386
Appendix: Auxiliary Overmarking Errors----------386

Chapter 4: The Grammar of SAI

4.0 Introduction----------394
4.1 Observations on SAI in English
4.10 Introduction----------394
4.11 Matrix Questions----------395
4.12 Tag Questions----------398
4.13 Comparatives----------400
4.14 So-preposing----------401
4.15 Subjunctives----------403
4.16 Negative Preposing----------404

4.17 Summary----------406

4.2 SAI and GB Theory

4.20 Introduction----------407

4.21 Safir's HUP----------408

4.22 Case-Based Theories of SAI----------413

4.23 Operator-Government Theories of SAI----------414

4.3 SAI as A'-Type Predication

4.30 Introduction----------426

4.31 A'-Type Predication----------427

4.32 P-Licensing----------429

4.33 Exceptional Application of SAI----------437

4.34 Exceptional Non-Application of SAI----------444

4.35 Conclusion----------462

4.4 SAI-Type Inversion Cross-Linguistically

4.40 Introduction----------462

4.41 General Cross-Linguistic Characteristics of SAI-Type Inversion----------464

4.42 "V2" Languages----------466

4.43 INFL2 in German----------467

4.44 Variation in INFL2-Related Phenomena in Germanic----480

4.45 Complex Inversion ----------505

4.46 Obligatory Inversion in Romance----------537

4.5 Principles and Parameters of SAI-Type Inversion

4.50 Introduction----------552

4.51 Parametric Variation and P-Licensing Condition II---554

4.52 The AGR-in-COMP Parameter----------567

4.53 The WH-Adjunction Parameter----------570

4.54 Possible P-Subjects----------573

4.55 The Dependent/Independent Clause Parameter----------575

4.56 Summary----------585

Chapter 5: The Acquisition of SAI-Type Inversion

5.0 Introduction----------593

5.1 What the Parameters Predict

5.10 Introduction----------594

5.11 P-Licensing Parameters----------595

5.12 Other Sources of Parametric Variation----------596

5.13 Summary----------598

5.2 Overview of the Acquisition of WH-Movement

5.20 Introduction----------599

5.21 Early Acquisition of WH-Phrases----------599

5.22 Acquisition of Adult-Like WH-Movement----------601

5.23 Lexical Specificity in the Acquisition of WH-Movement----------607

5.3 The Acquisition of SAI: Longtitudinal Studies

5.30 Introduction----------612

5.31 The Harvard Children----------612

5.32 The Child S.----------617

5.33 The Child K.----------622

5.34 Jessie----------644

5.4 The Acquisition of SAI: Cross-Sectional Studies

5.40 Introduction----------650

5.41 Ingram & Tyack (1979)----------651

5.42 Kuczaj & Brannick (1979)----------653

5.43 Erreich (1984)----------663

5.44 Stromswold & Pinker (1986)----------669

5.45 Summary----------670

5.5 The Acquisition of INFL2

5.50 Introduction----------680

5.51 Stages I & II----------681

5.52 Stages III & IV----------684

5.53 Summary----------687

5.6 Previous Accounts of the Acquisition of SAI

5.60 Introduction----------689

5.61 Bellugi's Account of SAI Development----------689

5.62 Klein's Account of SAI Development----------695

5.63 Pinker's Account of SAI Development----------703

5.64 Input-Based Acounts of SAI Development----------708

5.7 An Integrated Account of the Acquisition of SAI

5.70 Introduction----------714

5.71 SAI, INFL2, and AGR----------716

5.72 SAI Development and the Non-Inversion Stage---------723

5.73 The Interaction of Yes-No Questions and WH-Questions in the Acquisition of SAI----------726

5.74 Closed-Class Learning and the Acquisition of SAI----730

5.75 Residual Problems----------733

5.8 SAI Acquisition, Parameter-Setting, and the Two-Tiered Model of Syntactic Development

5.80 Introduction----------738

5.81 Linear-Parameter Setting and the Acquisition of SAI----------739

5.82 A Two-Tiered Model of Syntactic Development---------742

Chapter 6: Conclusion: On the Relation Between the Theory of Grammar and the Study of Language Acquisition

6.0 Introduction----------749

6.1 The Responsibility of Syntactic Theory to Language Acquisition

6.10 Introduction----------750

6.11 From "Instantaneous" to "Logical" Language Acquisition----------750

6.12 From "Logical" to "Plausible" Language Acquisition--755

6.2 The Responsibility of Language Acquisition to Syntactic Theory

6.20 Introduction----------758

6.21 Input Conditions----------760

6.22 Learning Mechanisms----------762

6.3 Conclusion----------763

References----------765

CHAPTER 1

INTRODUCTION

1.0 Preliminaries

1.00 Introductory Remarks

This thesis is an attempt to extend our understanding of the relation between the theory of grammar, conceived as an abstract representation of "what a child must know in order to learn a possible human language" and the theory of (first) language acquisition, by which we understand the specification of the mechanisms responsible for effecting the change from an initial "prelinguistic" state S_o to a final "adult" steady state S_s.

In this preliminary section, we will present - without detailed justifications - the metatheoretical underpinnings of this enterprise. Some of them are controversial, but we have no intention of entering this controversy from a philosophical standpoint; as Wittgenstein (1953) pointed out, such debate is useful only in "clearing away the rubble" in order to allow the constructive work of theory-building to continue. Ultimately, theories are not evaluated on the basis of their philosophical implications, but on other criteria, some of them internal (for example, redundancy and contradiction within the theoretical framework) some external (ranging from empirical success or

failure to purely institutional pressures); all these add up to the ability to "go on", to quote Wittgenstein again. The next seven hundred pages or so, whatever their intrinsic merit, testify to that ability.

We will adopt without further discussion the central tenets of what has become known as the "rationalist" or "Cartesian" view of the study of mental representations, amongst which, obviously, we include the study of grammar (for philosophical justification see Chomsky 1965, 1980b.,1986a.). According to this view, a grammar is a representation of a mental state, which is in turn a representation of a neurological state; correspondingly, "learning" a language involves a change, mediated by linguistic and extra-linguistic experience, from an initial (genetically endowed) state to a final steady state - the adult grammar.

For linguists working within this rationalist framework, the most crucial issues in the field revolve around the relation between the initial state (S_o) and linguistic experience; in particular - to pose the classical question - how does a child, given S_o and and a body of accessible linguistic evidence, come to "learn" the adult grammar (i.e., attain the steady state S_s) ?

In order to answer this question, clearly several others must be dealt with first, in particular, those concerning (a) the nature of the adult grammar (S_s); (b) the nature of the initial state (S_s); (c) the nature of the mediating evidence, usually known as Primary Linguistic Data (PLD); and (d) the

nature of the "learning" mechanism that effects the transition from S_o to S_s. Moreover, as a prerequisite to all of these questions, we must have an explicit definition of what it is to have "learnt" a language in the first place.

Obviously, some these questions permit a very wide range of answers. In the the next four sub-sections, we will introduce one particular approach to them - the paramweter-seting model which we shall adopt throughout this thesis. It is beyond the range of this study to explore all the possible alternatives; we merely hope to show that the particular line which we will take leads to some interesting conclusions, and generates some more interesting questions. And so we "go on".

1.01 Learnability and language acquisition

In this section, we will give a very cursory exposition of the most important - and relevant - issues in learnability theory - that is, the study of the conditions under which various types of mathematically defined languages are "learnable" - in a sense to be made precise immediately below. For a readable and thorough - though, inevitably, obsolete - review, see Pinker (1979).

Learnability is a very young sub-discipline - it dates from the publication of Gold (1967), approximately a decade after Chomsky (1957) showed how mathematically defined classes of language - those on the "Chomsky Heirarchy" - could be related to the study of natural languages. Gold defined language

learning in terms of a paradigm known as Identification in the limit; simply put, a language has been identified in the limit when no string from that language will force a learner to alter the grammar s/he has hypothesized on the basis of previous strings. Gold showed that given a completely general learner, certain classes of language in the Chomsky heirarchy were identifiable in the limit, depending on the way in which the strings were presented: if both grammatical and ungrammatical strings (appropriately labelled, of course) were presented, then the class of primitive recursive languages could be identified; on the other hand, if only grammatical strings were presented, only the finite cardinality languages were identifiable.

Now, the relevance of these findings to the acquisition of natural languages is this: (a) it appears that children only make use of grammatical strings (see 1.02 below) and (b) human languages are certainly of greater complexity than the finite cardinality languages (which simply consist of a finite list of strings). Gold's paradigm thus creates a paradox for any approach to natural language-learning based on a general learning procedure and classes of language in the Chomsky heirarchy, since, given these assumptions, human languages appear to be unlearnable.

In order to avoid "Gold's paradox", a number of moves are possible. The most important of these are (i) ordering the data-presentation; (ii) relaxing the identifiability criterion (iii) introducing a stochastic element into the learning

procedure; and (iv) constraining the learner's hypothesis space. It has been shown that any of these procedures will enable a learner to identify the class of primitive recursive languages (which certainly does include natural languages). The question then boils down to plausibility: given what we know about the actual process of human language acquisition, which of (i - iv) above appears to characterize the process most accurately ?

In fact, this question is a lot harder to answer than it at first might appear, since there is no a priori reason and little decisive empirical evidence to determine why children should not use any or all of these routes to language acquisition. Thus even (i), though implausible, cannot be discounted immediately; it is true that adults do not strictly order the complexity of sentences which they address to young children, but it is possible that the children themselves filter out much of their input and thus create the required ordering. As for (ii), it may well be that children do not identify languages in the limit, but simply converge on an approximation of the target grammar; for sentences of less than a given complexity, this type of learning will be equivalent to identification in the limit. Likewise, (iii) may appear initially implausible - or perhaps, aesthetically unpalatable - but it cannot be ruled out on those grounds, and in fact "learning by probability" has been proposed as a model of natural language acquisition by several researchers, most notably Labov and Labov (1978). And (iv), while the most widely

accepted solution to the learnability paradox, has been criticized on purely formal grounds (see for example Levelt 1979, who argues that even reducing the number of possible grammars in the child's hypothesis space to two makes no difference to decidability as long as the grammars are not entirely disjoint and are in the class of recursively enumerable languages).

Thus in many ways learnability theory tells us less about natural language acquisition than its initial results might suggest. What it does is to eliminate certain null hypotheses concerning language learning, and then throw the burden of explanation back onto empirical research; once we reach a certain level of theoretical sophistication, the question of human grammar learning cannot be answered except through the study of human grammars and children acquiring them.

Suppose, for example, we were to constrain the hypothesis-space of the language-learner so that the set of humanly possible languages is finite - this is Chomsky's (1981) position, which we will in fact be adopting as a working model. In that case, we avoid Gold's paradox, since language learning (at a certain level of abstraction) involves the choice between a finite set of possibilities; as long as the decision procedure which chooses between these possibilities is sufficiently stringent to ensure identifiability (and this may involve just one grammar with a set of parametrized variants), then the languages will be learnable. However, this is by no means the only way to ensure learnability, as we have just

seen, and moreover, there are many ways to restrict the learner's hypothesis space; we could, for example, restrict human languages to a certain class within the Chomsky heirarchy (the position of Gazdar 1982), or alternatively, claim that there need be no relationship between the Chomsky heirarchy and the set of possible human languages (the "scattered grammars" approach advocated by Chomsky 1981). The choice between these hypotheses is an empirical one, but this does not mean that it is simply a matter of finding the right kind of evidence or counter-evidence; rather, a great many considerations concerning both the internal and external dynamic of a given research paradigm enter into the interpretation of any given body of evidence. For example, the claim originally made by proponents of Generalized Phrase Structure Grammar (GPSG) that all natural languages could be described by context free grammars (which generate a particular class of languages in the Chomsky heirarchy) has been recently shown to be false: it is now known that at least two natural language constructions (cross-serial dependencies in Swiss German and reduplication in Bambara) cannot be described by context-free grammars (see Shieber 1985 and Culy 1985, respectively). However, this has not resulted in the abandonment of the GPSG enterprise, and it would have been very surprising if it had; the usual move in such circumstances is to minimally modify the original assertion (natural languages are now describable using only indexed grammars) but not the metatheoretical weltanschaung which originally inspired it (in the case of GPSG, this might

best be described as the conviction that formal rather than substantive universals are most likely to provide the basis for explanatory linguistic theories, and not coincidentally, computational implementations thereof - see Gazdar, Klein, Pullum and Sag 1985).

The point of this lengthy prolegomena is to emphasize that (a) linguistics is an essentially empirical discipline and (b) that it is not simply an empirical discipline, since empirical research in any even moderately advanced field of inquiry is a complicated business. These points are particularly relevant to attempts - such as this one - to integrate complex data with a wide variety of possible determinants into a particular research paradigm, since I do not even pretend to have produced the only possible, or "simplest", or even "most elegant" account of the "facts" themselves. Rather, I hope to show that under a particular interpretation of a particular theoretical framework, the data under consideration can (a) be accounted for and (b) can be accounted for in a way that sheds some interesting light on the "division of labour" between grammatical and non-grammatical factors in first language acquisition.

1.02 Universal Grammar, principles, and parameters

The particular research paradigm that we will be adopting in the rest of this chapter (and indeed, the rest of this thesis) is the "parameter-setting" model of language acquisition. There

are three main reasons for doing so. The first is that this model, as we shall see, can be made to yield an explicit set of predictions concerning the relation of language acquisition to the theory of grammar. The second is that it has generated a very productive and empirically interesting body of research, both in "pure" syntax and in the field of language acquisition. The third lies in its potential to unify the study of several hitherto distinct fields of linguistic and psycholinguistic inquiry, including those of comparative linguistics, historical linguistics, and real-time language acquisition.

In terms of the theory of language acquisition, the parameter-setting model provides a particular set of answers to the question we posed at the beginning of this section, concerning the relation between the child's initial state S_o - in learnability terms, the child's hypothesis space - and the steady state S_s, the adult grammar. These answers are cast in terms of what has become known as Universal Grammar (UG).

The theory of UG predicts that - given a certain level of abstraction - all individual human grammars will turn out to be closely related. This is not a new move - within the western grammatical tradition, it is at least as old as the 17th century Port Royal grammar - but within the Chomskyan paradigm, it does have a novel interpretation: UG is taken to be a specification of the innate species-specific genetic endowment that allows humans to learn any of the possible natural languages, and no others. In other words, UG = S_o, the initial state of the language learner, a highly constrained hypothesis

space from which the language learner may choose only a finite number of possible grammars.

Under this interpretation, the study of grammar becomes - indirectly, one must emphasize - the study of "what every child must know" in order to acquire a possible human grammar, given certain auxiliary factors (i.e., the role of the input and of the learning mechanism). Within the theory of grammar, this perspective has created a rich and productive research paradigm, based on extrapolating from the properties of individual grammars (i.e., a set of S_s's, or "final states" of the language learner) to UG (i.e, S_o, the initial state). In particular, in recent years the study of cross-linguistic universals (previously of merely statistical and historical interest) has gained renewed impetus, though on a more abstract and theory-internal level, as linguists attempt to describe the principles (which describe formal and substantive universals) and associated parameters (which define the limits of variation) of UG.

A principles-and-parameters model <u>is</u> in this sense a theory (on an idealized and abstract level) of language acquisition. The "logical" language-learner invoked so frequently in recent syntactic work is seen as approaching the task of acquiring a natural language heavily equipped by UG with a set of principles, which apply invariably to given syntactic subcomponents, and a set of parameters, which are "set" on the basis of PLD that is accessible to the child (so-called "triggering evidence"). The child, according to this view, does

not so much "learn" a language as choose one; correspondingly, the relationship between the learner and the PLD is indirect, rather than direct. It is to the question of the role of evidence in language acquisition to which we turn next.

1.03 The Role of Evidence in Language Acquisition

Clearly, the role of PLD in language acquisition is not negligible - if it were, then we would all be speaking the same language, and none of us would need to learn it. Some types of learning are obviously heavily dependent on the input, such as the acquisition of individual lexical items. On the other hand, given that UG restricts the range of grammars available in principle to the child, it is quite possible that the influence of PLD on grammar-development is severely limited; and indeed, the "poverty of stimulus" argument of Chomsky (1965) is based on exactly this premise. Chomsky's claim in "Aspects" is that the linguistic evidence presented to young children is both "noisy" (i.e., full of false-starts, slips of the tongue, and other speech errors) and inadequate (in that the types of sentence necessary for the linguist to decide between competing grammars are very unlikely to ever be presented to the child). To quote Chomsky himself:

> "Except for a ridiculously small number (e.g., conventionalized greetings, etc., which, in fact, often do not even observe grammatical rules) all actual sentences are of a probability so low as to be effectively zero, and the same is true of structures (if, by the structure of a sentence, we mean

the sequence of categories to which its successive words or morphemes belong). In actual speech, the highest probability must be assigned to broken and interrupted fragments of sentences which begin in one way and end in a different, totally incompatible way..." (Chomsky 1963, p.37)

It is fair to say that these claims are exaggerated, particularly in relation to speech to young children (see Snow 1986 and references therein); "Child Directed Speech" is generally slow, clearly articulated, and grammatical. However, the essence of Chomsky's argument remains intact, more than twenty years after it was first put forward: the PLD which constitutes the child's input severely undetermines the grammar which the child eventually constructs. In other words, the way in which a child constructs a grammar is qualitatively different from the methods used by linguists; the latter have access to all kinds of data which are inaccessible to the child, yet do a far less successful job (usually, over a much longer time period) than the average three year-old.

In fact, in the years since the poverty-of-stimulus argument was first put forward, it has received additional support from several quarters. One of the strongest of these additional arguments concerns the fact, initially reported by Brown and Hanlon (1970), that young children seem neither to receive nor to make use of direct negative evidence - that is, evidence in the form of overt corrections. This finding has rather profound implications for the theory of grammar (and for learnability theory - see 1.12 above), since it means that the child cannot use the ungrammaticality of a given sentence to draw inferences

about the grammar; this further sharpens the distinction we drew above between the child and the linguist as grammar learners, since - as any linguist will attest - ungrammatical sentences play a decisive role in most syntactic analyses. (Note, though, that the possibility remains that children might be employing some form of indirect negative evidence; for example, they might be calculating the probability of appearance of a certain form in a given context, and inferring ungrammaticality from its absence. See 1.12 above and 1.15 below for some comments).

In fact, there is other evidence of a more direct empirical nature supporting the claim that the child's use of PLD is severely limited: this is the frequently made observation that children acquiring a first language often appear to ignore easily accessible, frequently encountered evidence in constructing their initial grammars; in other words, grammatical development is orthogonal to the PLD. A prime example of this - which we will explore in great detail in Chapter 4 - is provided by children learning the rule of subject-auxiliary inversion (SAI) in English. Though (inverted) questions constitute the single most frequent sentence type in speech to young children (Newport, Gleitman, and Gleitman 1977), children at first fail to use any inversion at all (producing questions by using declaratives with question intonation) and even after learning the co-occurrence privileges of auxiliary elements, often fail to apply the (obligatory) inversion rule in WH-questions. This developmental

pattern cannot be explained on the basis of the child's input, even taking into account other variables such as "functional load", sentence stress, etc.; see Chapter 5 for details.

We conclude, then, that in its essentials the poverty-of-stimulus argument is correct: the role of the input in grammar-construction by children is severely limited, perhaps to the minimal role of a "trigger experience", as suggested by recent work in the "principles and parameters" framework. Correspondingly, the role of UG itself looms larger; the more richly specified the initial state S_o, the less evidence is needed for the child to reach S_s.

Given, then, that grammatical development proceeds orthogonally to the PLD, we still need to know what type of evidence can serve as triggering experience for the child. Since young children appear to have rather severe limitations on the length and complexity of utterances which they can process, it has been generally assumed that the relevant linguistic evidence must be present in short and relatively simple sentences. One way of encoding these pretheoretical notions is through the concept of degree of embedding, first used in the important work of Hamburger, Wexler and Culicover (see in particular Wexler and Culicover 1980). The degree of embedding of a node in a phrase-structure tree is calculated by adding up the number of "cyclic nodes" (NP and S) which dominate the node in question; if there are none, then the node is of degree-0 embedding, if one, then of degree-1 embedding, and so on. It is a particular virtue of the

Hamburger/Wexler/Culicover model that they are able to prove the learnability of a certain class of transformational grammars using degree-2 evidence only, thus providing a reasonably plausible model of the type of PLD that a child might be able to use in real-time language acquisition; however, at least for the early stages of grammatical development, even this may be too unrestricted a use of evidence, since young children appear to be able to process only very simple sentences, yet set important parameters. Recently, Culicover and Wilkins (1984) have in fact attempted to construct a theory of grammar based on the premise of degree-0 learnability; though this is clearly too restricted for certain parameters (e.g., the choice of "bounding node" for subjacency, as described by Rizzi 1982), it might be that at different stages of acquisition different degrees of evidence are available to the child, as processing capacities increase: this possibility will be kept open in our exploration of syntactic development.

Another input factor that has not been explored in the same depth as that of grammatical complexity is that of frequency in the input. We can distinguish two types of frequency, which we will label type-frequency and token-frequency. Type-frequency refers to the number of distinct instances of a given category (which could be a set of phonological features, or a set of grammatical features, or any other theoretically relevant class of items); token-frequency refers to the number of identical instances of a given item within a given category.

In fact, in our investigation of certain types of syntactic acquisition, we will argue that the role of frequency in the input, though important, is highly restricted. Essentially, we will be claiming that frequency effects are a characteristic of the learning of closed-class grammatical categories; this is unsurprising, since closed classes by definition contain a finite (in fact, small) number of items. We will show in addition that a number of puzzling aspects of data from language acquisition - including "lexical" effects and "token variability" - can be accounted for on the basis of the open/closed class distinction plus a certain type of learning mechanism (see 1.04 and 1.23 below).

1.04 Acquisition Mechanisms

Of all the components of the hypothetical model of language acquisition proposed in 1.00 above, the least-studied and the least well-known is the acquisition mechanism itself. Virtually all work in language acquisition, from the most mathematical abstraction to the most nose-to-the-ground empiricism have adopted virtually the same method of "learning", the analysis-by-synthesis or hypothesis-testing model. To illustrate the operation of this model, we will use the implementation presented by Wexler and Culicover (1980) as part of their demonstration of the learnability of a certain class of transformational grammars.

In the Wexler/Culicover model, the learner undergoes an

infinite series of trials which involve the presentation of a surface string together with its base-structure (which is available to the learner via the meaning of the string); these are known as (b,s) pairs. For each trial, the mapping from b to s is either compatible with the learner's transformational component (which effectuates the mapping) or it is not. In the latter case, the learner either rejects one of the (finite set of) transformational rules which comprise the current transformational component, or adds a new one (from the finite set which is a priori available). Given certain constraints on the type of transformations available, Wexler and Culicover show that this learning procedure will converge on the correct grammar for a given language.

While constructed for an abstract mathematical learning procedure, and thus not explicitly designed with real-time language acquisition in mind, the Wexler/Culicover learning mechanism is useful since, as so often with mathematical models, it provides an idealized null-hypothesis against which real-time acquisition can be assessed. It is thus worthwhile to detail the ways in which this procedure resembles, and the ways in which it differs from, the kind of learning mechanism necessary to account for the observed course of syntactic development.

On the plus side, the Wexler/Culicover model has two big advantages. Firstly, the learner need not employ negative evidence, since learning is driven by mismatches between (b,s) pairs; and second, there need be no "cross-checking" of data:

that is, the learner need not consult any previously stored sentences, since the grammar itself cumulatively encodes grammaticality decisions. Moreover, the PLD needed for the Wexler/Culicover model to learn a grammar is demonstrably limited to degree-2 complexity, that is, to sentences containing only two degrees of embedding; and in addition, the model can cope with a limited amount of "noise" in the data, since, given enough time, grammatical sentences will enable the learner to reconverge on the appropriate grammar.

The Wexler/Culicover model then does surprisingly well at mimicking real-time acquisition, a goal it was never designed to achieve in the first place. However, there are two ways in which it does not present a realistic model of child language acquisition. First of all, it is what Labov (1972) refers to as categorical; that is, it operates with transformational rules which either do or do not apply, but cannot apply variably. Secondly, it is insensitive to frequency in the input; it is as likely to reject a transformation on the basis of one token as on the basis of a thousand. As we shall see, neither of these properties characterize real-time language acquisition.

In order to overcome these problems, we shall propose the adoption of a slightly more sophisticated learning mechanism, a "parallel distributed processing model" of the type described by McClelland and Rumelhart (1986). It will be shown that by doing so we can avoid many of the problems associated with attempting to force grammars to do the work of learning mechanisms, and at the same time account - in a reasonably

principled way - for some characteristic properties of developing grammars, including both variability and frequency effects.

1.05 Summary

So far, we have provided (one set of) initial answers to the questions we posed at the beginning of this chapter, and can now provide a crude model of language acquisition, as follows: the child comes into the world equipped by heredity with UG, a set of principles and parameters which narrowly constrain his or her hypothesis space, and - together with a particular type of learning mechanism - enable him or her to learn a language (i.e., identify a grammar) on the basis of positive, accessible Primary Linguistic Data.

In the rest of this chapter, we will examine the implications of this model for real-time language acquisition. As we shall see, though the parameter-setting model holds much promise for a theory of syntactic development, it has proven far from easy to implement in the often confusing and confused world of child language.

1.1 Parameter-Setting

1.10 Introduction

While the parameter-setting model provides a fairly explicit

picture of what has been referred to as the "logical theory of language acquisition" (see for example, the papers in Hornstein and Lightfoot 1981 and Baker and McCarthy 1981) it makes no specific predictions as to how this "logical" theory should be implemented as a theory of real-time language acquisition, or, indeed, if it should be so implemented at all. In this section, we will examine several ways in which this might be accomplished, and briefly examine some of the work which has already been done in this area, before focussing in 1.3 on some pervasive problems which these initial attempts have encountered.

1.11 Ordering the Parameters: the Linear-Parametric Model

So far we have defined a parameter as a "choice" provided by UG which a language learner must make on the basis of appropriate PLD in order to generate the correct grammar for a given language. We have not yet, however, said anything about how many possibilities a parameter might encode (e.g., two or a thousand ?) or whether the possibilities are "weighted" (e.g., is possibility (a) likely to be hypothesized before possibility (b), or are they "equipollent" ?). In addition, though we have indicated that the relation of the parameter-setting language learner to the PLD is indirect, we have not indicated <u>how</u> indirect: is the role of evidence limited to a "checking" function at certain crucial points in language acquisition, or is the learning device more "powerful" and thus more sensitive

to the PLD ?

In fact, answers to these questions have been up to now rather surprisingly unanimous. The model which has emerged we will call the "linear-parametric" model, for reasons which will become obvious very shortly. Since virtually all work on the application of the parametric model to real-time language acquisition has been done within this framework, it constitutes a kind of "standard theory" for parameter-setting, one which, as we shall see, embodies both the major strengths and weaknesses of the approach.

A typical linear-parametric model consists of a set of syntactic principles with an associated set of strictly ordered parameters. There are three criteria which determine this ordering, one derived from learnability theory, the second from the theory of cross-linguistic markedness, and the third from "stages" in language acquisition. In the next three sub-sections, we will deal with them one at a time.

1.12 Parameters and Learnability

First of all, as shown by Berwick (1982) and Wexler and Manzini (1987), learnability considerations impose a certain intrinsic ordering on a set of parametric alternatives. The argument runs like this. To start with, it is assumed children have no access to negative evidence. Next, suppose a grammar G1 generates a language (i.e., an infinite set of strings) L1 which is a super-set of a language L2 generated by a grammar G2 - that is,

all the strings in L2 are in L1, but not vice-versa. Now, imagine a child faced with the (parametric) choice between G1 and G2. Suppose the child chooses G1 first, but the language to be learnt is in fact L2. In that case, the only evidence that could possibly tell him or her that G2 was in fact the correct grammar would be in the form of information that certain strings generated by G1 were ungrammatical in L2 - i.e., negative evidence. On the other hand, if the child were to choose G2 first, then positive evidence (in the form of strings generated by G1 but not G2) would be available in case the language to be learnt was L1, and obviously, there would be no problem if the language turned out to be L2. In other words, learnability theory imposes an ordering of parametric choices running from subset to superset: this principle is called the Subset Principle, following Berwick (1982).

It should be pointed out, however, that once we leave the intellectually hygienic world of learnability theory, the Subset Principle does not make as clear-cut predictions as might appear initially. There are a number of reasons for this, which we will now briefly examine.

First of all, in order for the subset principle to apply, the languages generated by two competing grammars must obviously be in a subset relation. However, this is not necessarily the case; it is quite conceivable that the languages might be partially or wholly disjoint. For example, in Hyams' (1983) treatment of the "pro-drop parameter" (which we will review in more detail in Chapter 2) languages with and

without pro-drop are not in a subset relation. This is because though pro-drop languages such as Italian have the option of generating "empty" subjects in tensed clauses, whereas non-pro-drop languages such as English do not, the latter type of language has overt pleonastic subjects (e.g., it, there) whereas the former does not. Therefore, the subset principle is irrelevant to the ordering of parametric alternatives for pro-drop. In order to ensure that the subset principle applies, Wexler and Manzini (1987) propose the following (pre)condition, which they call the Subset Condition:

(1) For every parameter p and every two values i, j, of p, the languages generated under the two values of the parameter are one a subset of the other, that is, L(p(i)) is a subset of or equal to L(p(j)), or L(p(j)) is a subset of or equal to L(p(i)).

(1) is an empirical claim, though motivated by formal considerations; thus Wexler and Manzini are committed to an alternative to Hyams' account of pro-drop, which is impossible given the subset condition. Of course, it is always possible that the subset principle is relevant for only some parameters, a weaker claim than that made by the subset condition. Again, this question can only be decided by empirical evidence, but it appears to me that parametrizations of such phenomena as the ordering of constituents within the phrase-structure component (see Williams 1981, Stowell 1981, Koopman 1984, Travis 1984)

which involve such parameters as "head right versus head left" are in principle incapable of being dealt with by the subset condition (see Safir 1987 for a similar conclusion).

A second problem attending the subset principle has been labelled the "many parameters problem". Suppose the setting of one parameter was contingent upon the setting of another; this is not a wildly improbable claim, as can be seen from the following example, provided by Williams (1987). It is widely accepted that a parameter of UG determines whether a language has overt syntactic WH-movement or not (English and Hausa do, Basque and Japanese do not). It is also widely accepted, following Rizzi (1982) that some kind of parameter must be associated with the principle of subjacency (usually formulated in terms of possible "bounding nodes", though see Adams 1984 for an alternative). However, since subjacency is itself a condition on (long-range) syntactic movement, it is quite obvious that the subjacency parameter is contingent on the setting of the WH-parameter. Of course, there are ways out of this (it might be claimed that in a language like Japanese the subjacency parameter was set "vacuously" at the most restricted setting: this would explain the bounded nature of scrambling, for example); nevertheless, the logic of the example is clear. To quote Williams, who puts it well, "...in the worst case, one can imagine the parameters were so paralyzingly interconnected that they all had to be set "at one time" and the evidence was the union of all the evidence relevant for any of them." In fact, such a scenario is not implausible, and even makes

predictions for the course of language acquisition: suppose, for example, that children remained in a "pre-parametric" state until the relevant evidence had been encountered, and then, setting every parameter at once, emerged butterfly-like from the cocoon of evidence that they had woven. The only additional assumption we would have to make would be that children were capable of "storing" certain types of crucial information - and thus, capable of knowing what types of evidence to store -prior to applying them at the moment of "emergence"; as Williams points out, this is precisely the assumption made by proponents of the "evaluation metric" approach to grammar-determination.

Once again, Wexler and Manzini (1987) are forced to construct an auxiliary principle in order to maintain the relevance of the subset principle and to eliminate the type of scenario outlined above. This time, it is the principle of Independence:

(2) The subset relations between languages generated under different values of a parameter remain constant whatever the values of the other parameters are taken to be.

It is not at all clear to me that (2) has any empirical content, for the following reason. Take a parameter such as the pro-drop parameter, as originally formulated by Rizzi (1982), Jaeggli (1982), and Chomsky (1981). At that time, it was believed that a single parameter could be responsible for a

large range of syntactic effects, including, in the case of the pro-drop parameter, "free" inversion, "null" subjects, "empty" expletives, and absence of a *[COMP t] effect. Moreover, it was accounted a particular virtue of the parametric model that superficially diverse syntactic effects could all be handled by a single "deep" parameter. Now, this raises the following problem: it is conceivable (indeed preferable, given the logic of the parametric approach, as exemplified by pro-drop) to treat any set of parametrically interconnected phenomena as evidence for a single parameter. In other words, there is no a priori way to distinguish between one parameter with diverse effects and many connected parameters; but if not, then the "many-parameter problem" can always be solved by broadening the scope of a single parameter. Of course, empirical considerations might lead to the discovery of diagnostics for telling parameters apart: it might be, for example, that separate but interconnected parameters are set at different times during the course of language acquisition (since they would depend on partially disjoint PLD), or that the effects of interconnected parameters had a different cross-linguistic distribution than those of a single parameter (since they might be implicationally related in only one direction) - however, at this point, no such diagnostics exist, as far as I am aware of. If not, then the independence principle as formulated in (2) makes no predictions at all unless supplemented with additional conditions on the possible "scope" of a single parameter.

Wexler and Manzini, following Borer (1984), do indeed come

up with a condition of this type, the Lexical Parametrization Hypothesis (LPH), as formulated in (3):

(3) Values of a parameter are associated not with particular languages, but with particular lexical items in a language.

It is unclear what effect, if any, the LPH would have on the type of "manifold" parameter exemplified by the Rizzi/Chomsky pro-drop parameter, which depends on the level of application of a morpho-syntactic rule, "Rule R". It is conceivable that it might be accomodated by claiming that "Rule R" applied to particular lexical items, i.e., instantiations of AGR; it is also conceivable that manifold parameters of the Rizzi/Jaeggli/Chomsky type do not actually exist (after all, the original formulation of the pro-drop parameter has passed into history). However, the LPH, while potentially solving the problems caused by the Independence Principle (which in turn, solves problems caused by the subset principle) leads to a new non-trivial problem, which we will refer to, essentially following Safir (1987), as the "undergeneralization" problem.

The problem boils down to this: given that parameters must be set separately for each lexical item, why don't we find entirely random scattering of values for a given parameter throughout the lexicon of a given language ? Take, for example, the binding theory (the very example that is used to motivate the LPH): while it is true, as Manzini and Wexler (1984) show, that in Icelandic and Italian the governing category for

different lexical items must be set separately, in English, all pronouns and anaphors appear to have the same governing category. This is surely not the accidental result of a whole set of lexical parameters being set the same way; but if not, then yet another principle (perhaps referring to a low-level "generalization tendency") must be invoked to account for the fact that in the unmarked case, individual lexical items group together into identifiable subclasses which all behave identically with respect to a given parameter. In other words, in attempting to account for the "lowest common denominator" of lexically determined syntactic behaviour, the LPH misses significant supra-lexical generalizations.

Thus we see that the logic of learnability, as embodied by the subset principle, which at first appears to promise an independent, non-arbitrary ordering of parametric variants, cannot be imposed on natural language acquisition without the additional imposition of a whole set of auxiliary principles, some of them empirically rather suspect. Once again, it appears that learnability theory can only guide us as far as the margins of the empirical wilderness.

1.13 Parameters and Cross-Linguistic Markedness

Let us then turn to our second putative approach to the ordering of parameters, this time based on comparative evidence, and embodied in the idea that the theory of markedness derived from the study of cross-linguistic variation

will reflect either directly or indirectly the ordering of parameters in language acquisition.

Clearly, in considering cross-linguistic evidence, we must distinguish "deep" from "surface" markedness. The latter will characterize such statistical surveys as that of Greenberg (1963), which though extremely valuable as a kind of concise cross-linguistic encyclopedia, bears no direct relationship to any particular theoretical framework, and cannot therefore bear on the question of parameter-setting. In contrast, the recent upsurge in work on cross-linguistic work specifically inspired by the parameter-setting model (see, amongst many others, Huang 1982, Borer 1984, Koopman 1984, Travis 1984, Safir 1985) is obviously of potential relevance.

Note, however, this kind of work is still only potentially relevant; even syntactic research cast overtly cast within a "parametric" framework does not necessarily bear directly on the implementation of that framework in real-time language acquisition. This point is made by Rizzi (1986), who suggests that "markedness" in acquisition is not necessarily a function of markedness between grammars - as cross-linguistic markedness is commonly understood - but may involve other non-grammatical factors specific to the acquisition process itself. In that case, we would not necessarily expect there to be a direct relationship between "cross-linguistic markedness" and "acquisitional markedness". A case in point is the pro-drop parameter, as applied to acquisition data by Hyams (1983). Rizzi argues that from a cross-linguistic perspective, the

"null subject option" is in fact the marked value of the parameter; yet Hyams' study purports to show that it is the unmarked option for the language learner.

Lebeaux (1987), following the analysis of Guilfoyle (1984), suggests that this paradox might be overcome by dropping the widely held assumption, first made explicit in White (1982), that every grammar hypothesized by the child must be a potential adult grammar; in that case, children might pass through grammatical stages which could not correspond to any attested cross-linguistic parameter-setting. Lebeaux goes on to suggest that some "complexity measure" might be devised to intrinsically rather than extrinsically order the child's developing grammars, thus giving content to the intuitive notion that grammars increase in complexity over time.

It seems to me, however, that Lebeaux's analysis gives up too much: in fact, virtually all of the more attractive features of the parameter-setting model. Most importantly, by abandoning the "possible-grammar" approach to childrens' language, Lebeaux is virtually forced to give up what is known (following Pinker 1984) as the continuity hypothesis. The continuity hypothesis essentially claims that the learning procedure - more specifically, the component of the learning procedure which includes UG - does not change during the course of language acquisition (apart from possibly "withering away" at the end of language acquisition, in a similar way to the Marxist conception of the State). Now, the continuity hypothesis is by no means necessarily true - indeed, Borer and

Wexler (1987) argue that it isn't supported by the evidence, either - but it has proved to be an extremely useful heuristic for grammar-based explanations of language acquisition, in so far as it imposes non-trivial constraints on the relation between "child" and "adult" grammars. In other words, we should not abandon it without a fight, for by doing so we allow a far less constrained theory of language acquisition, where the maturation of syntactic principles can be used as a deus ex machina to explain any given set of data from child language. Of course, it might be that a maturational theory is the only theory that can account for certain aspects of syntactic development - in particular, those where the the child appears to ignore relevant evidence that is accessible and easily processed - but it seems to me this is a worst-case scenario, to be adopted only when all else fails.

In fact, Weinberg (1987) offers yet another account of the acquisition of the pro-drop parameter (unfortunately, based once again on Hyams' less than convincing data - see Chapter 2 for a more detailed discussion) which dispenses with the need for an "impossible" intermediate grammar, and instead argues that a performance limitation interacts with the pro-drop parameter to force the child to adopt the more marked option. The performance limitation in question is responsible for the "two-word" stage in language acquisition, where although capable of understanding more complex utterances, the child is limited to the production of just two words. In intransitive sentences, this poses no problem, and indeed Weinberg cites

Goldin-Meadow (1979) in support of her claim that (overt) subjects are statistically correlated with the use of intransitive verbs in the language of children at this stage. However, when faced with having to produce a transitive sentence, a child at the two-word stage must either drop the subject or the direct object. Since it is assumed that the Extended Projection Principle of Chomsky (1982) holds for the relevant stage - of course, an assumption based on the continuity hypothesis - the child can maintain grammaticality in the face of the two-word constraint only by adopting the positive (marked) value of the pro-drop parameter. In that case, of course, the child will produce two-word utterances consisting of verb and direct object. In this way, the continuity hypothesis can be maintained and the "backwards parameter-setting" explained without recourse to more drastic measures.

It may be, of course, that not every putative mismatch between cross-linguistic and acquisitional markedness can be explained away in this manner. One example which seems less amenable to reinterpretation is the case of preposition-stranding. It has been known for some time that preposition stranding, as exemplified in (4a.) below, is cross-linguistically marked relative to pied-piping, the alternative in (4b.):

(4)a. What country do you live in ?

b. In what country do you live ?

Various explanations have been forward to explain this observation: see in particular van Riemsdijk (1978), Hornstein and Weinberg (1981) and Kayne (1984); all of them, however, share the premise that some special mechanism is needed to account for cases like (4a.).

Now, if acquisitional markedness were directly tied to cross-linguistic markedness, then we would expect children to start out by producing sentences such as (4b.); however, exactly the opposite is the case; it is literally years after they first employ structures such as (4a.) - if at all - that children learning English begin to use pied-piping structures such as (4b.). For experimental validation, see French (1984). Moreover, unlike in the pro-drop case, there is no immediately available performance explanation (of course, one could always be cooked up, but it would lack independent motivation). Note in addition that children who fail to pied-pipe have no problem with fronted prepositional phrases, or the use of "inherently" prepositional WH-phrases such as where.

Note that we cannot adopt a maturational explanation along the lines of Borer and Wexler, either. There is nothing to mature. The fact is that in English (of the adult as well as the child variety) pied piping is marked relative to preposition stranding.

In fact, there is an obvious reason why children learning English should acquire stranded prepositions so early and pied-piping so late: they are simply learning what they get in the input. In fact, I'd like to suggest that this is anything but

abnormal: in the vast majority of cases, the PLD available to a child at a given stage of acquisition, together with a reasonably robust learning mechanism and a sufficiently restricted set of parameters, is quite adequate to set a parameter the right way the first time. That is, rather than "blindly" running through a set of extrinsically ordered parametric alternatives from least marked to most marked, the child in general makes use of salient evidence as soon as it is available.

Now, this might seem counter to all that we have been saying concerning the relative "data-independence" of the parametric learner. However, here we make a rather specific prediction: if a child either undergeneralizes (i.e., fails to make use of the right evidence) or overgeneralizes (i.e, uses the wrong evidence) it is because the PLD itself is misleading. This in turn leads to the concept of acquisitional markedness represented informally in (5):

(5) A language L_i is acquisitionally marked just in case the subset of strings in L_i ($L_j(L_i)$), where L_j constitutes the PLD available to a child learning Li at time t, leads a child to formulate a grammar G_j which differs from the target grammar G_i at time t.

The concept of acquisitional markedness presented above is not meant to supplant the "intrinsic" markedness relations which characterize cross-linguistic variation. Neither is it meant to

imply that there is no relation between language acquisition and intrinsic markedness; this relation is probably the cornerstone of the parametric edifice, and as such I have no intention of kicking it out from underneath us and bringing the whole building down on our heads. Rather, what I am suggesting is that parametric learning is a good deal more efficient - and more instantaneous - than is currently envisioned, and it is only in the cases where the system "goes wrong" (i.e., where the PLD at a given stage of development misleads the learner) that the child will produce a grammar which is not equivalent to that of the adult.

One of the main advantages of this perspective is that it re-establishes an important connection between "misprojections" in language acquisition and the direction of language change, a connection first made by Halle (1962). The concept of acquisitional markedness elucidated above enables us to make the prediction that the child's misprojections will ultimately shape the direction of syntactic change - in this sense, an acquisitionally marked construction is a historically marked construction. It should be clear that the notion of intrinsic (cross-linguistic) markedness usually associated with the parametric model cannot make such a prediction: for example, while preposition stranding is intrinsically marked, there is no indication that it is disappearing in English, any more than there is evidence that it is hard to acquire.

Thus we have proposed a rather different acquisition model from those we have examined so far. In particular, it makes

different predictions concerning the course of real-time language acquisition, and the types of "intermediate grammars" which we will discover there. It is to these predictions that we now turn.

1.14 Parameters and Real-Time Language Acquisition

Let us review once more the mechanics of the "linear-parametric" model of language acquisition. According to this view of parameter-setting, the learner runs through a set of parametric variants running from least to most marked (however markedness is calculated); the role of the PLD is limited to a a "triggering function" activating the parameter and a "checking function" as each alternative grammar is matched against a set of PLD.

Now, it is obvious that, at least in the simplest case, such a model makes strong predictions concerning the course of real-time language acquisition. In effect, if a language sets a parameter at the least marked value, then we should expect the child to have no difficulty in acquiring it; on the other hand, if the language sets a parameter at a more marked value, then we should expect the child to pass through a series of "stages" corresponding to the less-marked values of the parameter.

Now, the question immediately arises as to whether such "stages" exist in first language acquisition. Indeed, a good part of this thesis will be concerned with exactly this question; we will be examining three well-known "stages" in the

acquisition of English: the "pro-drop" stage in Chapter 2, the "auxiliary overmarking" stage in Chapter 3, and the "non-inverting stage" in Chapter 5. We will argue that "stage-like" effects in language acquisition are not in fact the result of a linear-parametric type of language acquisition, but are genuine misprojections of the type described in the last section.

It should be noted, however, that even if children do not show stage-like acquisition sequences, there are a number of ways in which a proponent of the linear-parametric model might accomodate the data. It could be claimed, for example, that much acquisition was "silent" - that is, a child might pass through stages rather rapidly and efficiently, and show no overt evidence for linear-parameter setting. This is not an unreasonable hypothesis - in fact, we ourselves are claiming something very similar. The difference between our position and that of say, Hyams (1983) is that the latter is committed to an explanation for <u>overt</u> sequences in language acquisition based on the linear-parametric model, whereas we have claimed that there are no "stages", and that the observed sequences are to be explained as misprojections where, in effect, the child is led astray by the PLD.

The question, of course, is an empirical one, and cannot be answered in this brief introduction. We will therefore defer an answer until the end of the thesis, when we will have hoped to show that in at least three important cases, what appear to be "stages" in acquisition are in fact misprojections, supporting

our misprojection hypothesis over that of the linear-parametric model.

We will, however, introduce in the next two sections two rather central considerations in the analysis of child language data which bear directly on the question of how a syntactic theory should best be related to a theory of language acquisition. The first concerns variability, and the second performance factors in acquisition.

1.2 Variability

1.20 Introduction

In this section, we will examine one of the most controversial and difficult areas in the analysis of language acquisition data, the question of variability. Initially, let us distinguish between two major types of variability, inter-subject variability and intra-subject variability. The former refers to variation between children, the latter to variable production by the same child. In the next sub-sections, we will discuss these kinds of variability one by one.

1.21 Inter-Subject Variability

It is by now well-known in the field of child language acquisition that children show individual variation over a wide range of different developmental factors. First of all, there

is considerable variation in "rate of development" as calculated on the basis of such global measures as chronological age, Mean Length of Utterance (Brown 1973) and vocabulary size (Nelson 1973); for example, in the large scale longtitudinal Bristol Study of Language Development Wells (1985) found that the age-range difference between the "slowest" and "fastest" children during the course of "normal" language acquisition was 30-36 months.

More importantly for our purposes, there also appear to be significant individual differences in the "route" of language acquisition. As we shall see, even the most pervasive types of syntactic undergeneration and overgeneration are by no means universally attested in children's language at a given point in development; indeed, we will use this fact to argue against a linear-parametric approach to such misprojections.

What is the source of such individual variation ? Let us begin by saying what we believe it is <u>not</u>. In particular, we do not countenance (non-parametric) variation in UG itself; though this type of variation is not a priori impossible, the assumption that the fundamental principles of UG are the same for each child seems to me to be both a good working hypothesis and a necessary analytical heuristic given our current state of knowledge (or ignorance) of the relation between grammatical theory and language acquisition.

Given, then, that UG is not subject to variation, it appears that there are at least three (and in all probability, many more) determinants of variation. Let us at least distinguish

the following:

(a) The input. It is possible that differing inputs might lead to different grammars. Not much research has actually been devoted to this topic; what we know is primarily based on various interpretations of the Harvard Study of language acquisition (see in particular Brown and Hanlon 1970, Moerk 1980, 1981, Pinker 1981). Results are inconclusive (or rather, conclusive but diametrically opposed, depending on which analysis you attend to). One interesting study (Fee 1980) which attempted to trace the relationship between children's acquisition of pronouns and parental "pronominal styles" found no significant correlation between the two.

(b) "Style". Several studies have been devoted to the question of individual styles of language acquisition. On a very basic level, anybody who has ever done "field work" (actually, usually floor-work) with young children has noticed that some children are linguistically "bold" - even rash - while others are generally cautious; the former type are generally more illuminating for linguists. See Kuczaj and Maratsos (1983) for some apposite comments.

Other aspects of syntax-related individual style which have been investigated include the proportion of nominal to verbal elements in early language, nominal versus pronominal preferences, extent of use of unanalyzed "formulaic" strings, extent of spontaneous imitation, and "expressive" versus

"referential" language. For an extensive review, see Wells (1986) and references therein. It seems likely that as our knowledge of inter-subject variation improves, this list will increase rather than decrease - for example, children differ in the kinds of questions they ask, the way in which they make requests (direct or indirect), and so on.

(c) Performance limitations. Very little is known about the differential effects of various performance-related constraints. It seems highly probable, however, that these interact in non-trivial ways with developing grammars. An example given above is Weinberg's (1987) reanalysis of the data presented in Hyams (1983), where a performance factor forces the setting of a syntactic parameter the "wrong" way. Now, it seems likely that such factors - which are by definition grammar-independent - do not affect each child in the same way at the same developmental point; it is easy to imagine two children at the same stage of syntactic development reaching two rather different conclusions on the basis of an identical set of PLD, if one child has a severe short-term memory constraint, for example, which the other lacks. Borer and Wexler (1987) point out that it has somehow been easier for language acquisition theorists to imagine performance factors changing over time (and at different rates) than it has been to imagine UG changing over time (and at different rates), and claim that this is somehow intellectually inconsistent; however, in fact, this position seems to be the most sensible

approach to take at present, since (i) we know that performance factors do change over time (MLU, for example, is not an index of syntactic development, since a child could conceivably produce long strings of coordinated simple sentences without ever needing complex syntax; yet MLU changes over time, and the most reasonable hypothesis for this change is an increase in performance capacity); and (ii), for reasons argued above, it seems more sensible at present to hold UG constant while varying performance constraints than to assume that both UG and performance factors vary independently.

Thus we see that inter-subject variability is a complex and difficult issue, with a wide variety of almost certainly interacting variables at work (or play). Fortunately, we do not have to attempt to pick them apart; we simply point out here that we do not need to "relativize" UG or the parameter-setting model in order to account for the observed variability - if anything, there are so many available explanations that the data severely undetermines any particular analysis. For a different view, see Borer and Wexler (1987).

1.22 Intra-Subject Variability

Let us then turn to our second major type of variation, which we have labelled "intra-subject variability": that is, variable production by the same child at the same time. In fact, we will further differentiate intra-subject variability into type-

variability (where a child applies a rule variably according to lexical type) and token variability (where a child applies a rule variably to tokens of the same lexical type), since, as we shall see, these two sorts of variability have rather different implications for grammar-based theories of language acquisition. Accordingly, let us deal with them one at a time.

1.23 Type Variability

It has often been remarked (at least since Brown 1973) that, ceteris paribus, a grammatical (rule-based) approach to language acquisition predicts that the learning of a rule should apply to a given category "all at once"; for example, we would expect, if the rule of Subject Auxiliary Inversion (SAI) applies to all the members of the category "AUX", that a child should learn the rule simultaneously for all auxiliary elements. However, even the most cursory inspection of data from real-time language acquisition reveals that this prediction is manifestly false for almost all aspects of grammatical development (and also, it appears, for morphological and phonological development). Instead, we observe a similar pattern of acquisition over and over again, across different components of the system and at different points in development: at first, distributional regularities (i.e., rules) are acquired lexical item by lexical item, with little categorial generalization; there then follows an (often brief and sometimes missing) period of "over-regularization",

and finally the child settles on something approaching the adult system. This is the "U-shaped learning curve" made famous by studies of the acquisition of inflectional morphology, but which appears equally applicable, as we shall see, to many aspects of syntactic development: for example, this is exactly how children appear to learn SAI (we refer the reader to Chapter 5, where the acquisition of SAI is discussed in detail).

There are essentially three ways of accomodating this U-shaped learning pattern to grammar-based theories of first language acquisition: we can attempt to account for it through properties of UG itself; we can attribute it the influence of performance factors; or we can attribute it to properties of the learning mechanism. Let us examine these options one by one.

First of all, let us examine the possibility that the way in which children learn syntactic rules is a direct reflection of the way in which UG organizes those rules. In many ways, this is an optimal solution from the point of view of the connection between linguistic and psycholinguistic evidence, a point made particularly frequently by proponents of the so-called "psychologically real" grammatical theories (see Maratsos 1978, for example).

What kind of grammar, then, would account for the attested learning pattern ? To start with, it would have to be lexically-based. This is a property of several current syntactic theories, most notably that of Lexical Functional

Grammar (see the papers in Bresnan 1982 and in particular Pinker 1984), but also of some varieties of Government Binding theory (e.g., the "lexical" parametric model of Borer 1984 and Wexler and Manzini 1987). The main advantage of a lexically-based model is that it accounts immediately for the "lexical learning" patern which appears to characterize the acquisition of syntactic rules; if such rules are tied to specific lexical items, then we immediately explain why they are learnt item by item. An additional advantage is that a lexical learner will never encounter the "projection problem" first descibed by Baker (1979), in which an overgeneral rule-learner "goes too far" and is unable to retreat without the use of negative evidence. To put it bluntly, a lexical learner will never go too far.

These advantages, however, are not gained without a cost. As we have already pointed out, syntactic rules apply generally; exceptions <u>are</u> exceptions, not the rule. This is entirely mysterious within a strictly lexically-based theory, since - without the addition of auxiliary principles - a grammar in which each lexical item arbitrarily does or does not undergo a particular rule will be as highly-valued as one in which all lexical items of a particular syntactic category behave alike for a given rule. Thus lexically-based theories suffer from the "undergeneralization problem" discussed in 1.23 above.

Moreover, lexical learning is in many ways <u>too</u> successful as an acquisition device. It is true that it avoids Baker's projection problem; but it is also true that children often do

not. In the typical "U-shaped learning" case, undergeneralization is followed by overgeneralization; however one attempts to solve the learnability puzzle created by the ability of children to retreat from an overgeneral rule, the fact that they advance too far in the first place argues that lexical learning cannot be the right model for the real-time language learner.

We tentatively conclude that attempting to adapt UG to account for learning patterns is not a very promising move; this conclusion will receive additional support when we come to examine token-variability in children's language.

The next possibility is that performance factors are responsible for the observed pattern of rule-learning. The trouble with this is that it is very difficult to see exactly which performance factors might be responsible. In order to account for the U-shaped learning curve, the requisite performance limitation(s) would (a) have to explain initial item-by-item acquisition (b) subsequent overgeneralization and (c) final retreat. I know of no single possible performance-based explanation for such a sequence: see 1.4 below for a brief review of what is (and perhaps more importantly, what is not) known about performance factors in language acquisition.

This leaves one remaining possibility, that it is the learning mechanism which acounts for the observed pattern of development. I would like to argue that a solution along these lines is in fact the most promising line of inquiry. More particularly, I believe that a learning mechanism based on the

parallel distributed processing (PDP) model presented by McLelland and Rumelhart (1986) has exactly the right properties to account for the U-shaped learning curve so prominent in the learning of syntactic rules. In order to see why, we will briefly describe the salient properties of this model; more detailed descriptions can be found in McLelland and Rumelhart (op. cit.).

Essentially, a PDP consists of three very simple parts. The first, the input pool, consists of a representation of the input to a given rule; the second, the output pool, correspondingly consists of a representation of the output generated by the mechanism as its "current" guess. In between, there is a pattern association network consisting of a series of modifiable connections linking each input to each output, which may be initially "weighted", or not; if not, each connection will be initially set at 0. The machine "learns" by modifying the strengths of these connections through "experience" in the form of pairs of inputs and outputs.

Thus, when presented with a given input, the machine will activate the input units associated with the input-representation, then for each output-representation compute the net input from all of the weighted connections from the input units. The net input is the sum over all input units of the input activation times the corresponding weight. If the net input exceeds a given threshold, then an output unit will be turned on - i.e, the machine will make a "guess". In addition, a stochastic element is introduced, in the form of

a logistic probability function, so that the machine can make different "guesses" for the same input/output pair at different times.

Now, if the output representation generated by the machine matches the output given to the machine in a "learning trial", then the weights on the association lines will not be altered. On the other hand, if the guess is "wrong", then either the input weights are increased by a fixed amount and the threshold decreased (in the case of a "miss", where an association is not activated, corresponding to an undergeneralization) or the input weights are reduced and the threshold increased (in the case of a "false alarm", where an association is incorrectly activated, corresponding to an overgeneralization). It is known that such a procedure, called a Perceptron Convergence Procedure, will eventually find a set of weights that will perform correctly, as long as one exists.

As McLelland and Rumelhart show, this type of learning mechanism can be made to mimic the U-shaped learning curve for a given input rather successfully. Their example is the classic case of the English past tense, a well-known case of U-shaped learning (see Bybee and Slobin 1982). On a computer simulation incorporating a PDP learning mechanism, the learner was presented with the root and past tense forms of the 10 most frequent English verbs (including 8 with an irregular and 2 with a regular past tense) for a total of 10 learning trials, then presented with 410 medium frequency verbs (including 334 regular and 76 irregular past tenses) for 190 learning trials.

Finally, 86 low-frequency verbs (72 regular, 14 irregular) were presented to the system to test its response-pattern.

The results of this procedure show a familiar U-shaped curve: at first, irregular and regular forms are learnt in the same way; then, when the medium frequency verbs are introduced, performance on the irregular verbs deteriorates - in other words, the system "overgeneralizes" from the regular to the irregular cases, as a function of comparative frequency. Over time, performance on the iregular forms improves, but never reaches the level of regular verbs. McLelland and Rumelhart show that not only the general learning curve but the actual "over-regularizations" made by the system correspond fairly closely to the actual course of past-tense acquisition in English: see McLelland and Rumelhart (op. cit.) for details.

Now, while it should be noted that there are certain rather unnatural assumptions in this simulation procedure - in particular, the machine is presented with data in a particular (idealized) order - the results seem to me to be suggestive in a number of ways. First of all, we need no retreat mechanism for overgeneralization, since "overgeneralizations" are merely a particular stage in the learning procedure. Second, we account for lexical effects, since the learning procedure is lexically based. Third, we account not just for type-variability, but for token variability as well (see 1.24 below). And last, but not least, we relieve the grammar itself of having to account for both type- and token-variability, which are now described by features of the learning mechanism, not UG.

It should be emphasized that this is not the way McLelland and Rumelhart see things. Their claim is that associative network models such as the one just described can supplant explicit grammatical rules. On the other hand, the approach I will take here is that such networks are part of the learning mechanism, not the "grammatical knowledge component"; as such, they supplement explicit grammatical rules. To put it another way, grammar learning is a particular type of linguistic performance, and as such, learning mechanisms are performance mechanisms. What we have done by increasing (only slightly) the power of the learning mechanism is then to relieve the competence component of the task of having to deal with certain performance-based effects in language acquisition, a move which I hope will clarify the relative contributions of UG and the learning mechanism in real-time language acquisition.

1.24 Token Variability

While there have been various attempts to account for type-variability in language acquisition (some of which are described in the previous section), token-variability, i.e, the variable production of the same token by the same child at the same time, has received considerably less attention. This is not surprising, since token-variability presents a formidable problem to any grammar-based explanation of linguistic development. Nevertheless, the phenomenon is pervasive (if not always acknowledged) in the production of children at every

stage of syntactic development. By way of illustration, I present the following examples of WH-questions from K., a child whose development I followed during an eight month period from approximately 33-41 months (and whose grammatical system is described in detail in Chapter 5). All of the examples were produced during the same session (actually, within the same 30 seconds), when K. was 32 months old:

(6)a. What's he saying ?

b. What he saying ?

c. What he's saying ?

How do we go about explaining this type of variability ? As with type-variability, three possible explanations come to mind.

The first possibility is to build the variability right into the grammar, weighting syntactic rules with given probabilities. This is quite compatible with learnability considerations since it is known by Bayes's Theorem that a stochastic language learner, who in fact uses a type of indirect negative evidence, will be able to identify the class of primitive recursive languages, which includes natural languages (see Horning 1969). However, "stochastic learning" runs counter to most current grammatical thinking, perhaps mostly because it involves the unappealing premise that children come into the world equipped not only with innate grammatical principles but grammatical principles with

probabilities attached (though see Chomsky 1965 for a proposal for a grammar evaluation metric with almost exactly these properties). Perhaps the best-known proponent of this approach is William Labov; we shall examine his treatment of one set of language acquisition data in Chapter 5.

The second approach to the token-variability problem in language acquisition - one very popular amongst theorists - has been to discount it as "interference" from performance factors. The problem with this approach, as observed above, is that while very effective when little or nothing was known about the exact nature of such performance factors, it has become less and less plausible, the more we have discovered since. Essentially, this is an argument ad ignorantiem which, while unfalsifiable as long as there remain factors in language acquisition beyond our current knowledge, completely lacks explanatory potential, and therefore deserves to be consigned to oblivion.

This leaves us once more with an explanation based on properties of the learning mechanism. As we have seen, the McLelland/Rumelhart model has exactly the right characteristics to deal with token-variability; this is one of the strongest arguments in its favour. By building a stochastic element into the learning mechanism, a move which is independently necessary in order to deal with "noise" in the input, we avoid having to build a stochastic grammar, and all the difficulties which that inevitably creates (in particular, the abandonment of the "no-negative evidence" hypothesis, which has proved such a useful

heuristic in theoretical approaches to language acquisition). Moreover, we have no need to appeal to shadowy "performance factors" with no empirical basis. In other words, token-variability argues strongly for a learning mechanism such as the McCleland/Rumelhat model which we have adopted.

1.25 Conclusion

In this section, we have examined the issue of variability in language acquisition, a topic usually shunned by those attempting to apply theoretical principles to data from child language. We have shown that in fact, far from being unilluminating, variability sheds some interesting light on certain "demarcation questions" in the theory of language acquisition; in particular, by adopting an enriched learning mechanism, we can both account for the types of variability actualy observed during syntactic development and relieve UG of the burden of having to explain them.

We now turn to a murkier area, the question of the influence of performance constraints on language acquisition.

1.3 Performance Constraints and First Language Acquisition

1.30 Introduction

In this section, we will try to come to grips with the influence of "performance factors" on first language

acquisition. As the reader might have gathered by now, we are not entirely comfortable with performance-based explanations for child language data, for the simple reason that in most cases such "explanations" do not specify what the performance factors actually are. The fact is, in most cases we know considerably less about performance than about competence, though there tends to be an unconscious bias amongst psycholinguists to treat the former as "less abstract" and therefore a priori preferable as an explanatory mode. This seems to me to be quite wrong: in most cases, performance-based explanations must be at least as complex as competence-based explanations, since they include the latter. Take as an example the theory of sentence-parsing, a prototypical case of performance. Such a theory must include a competence component, unless one takes the strange view of Fodor, Bever and Garrett (1974) that parsing heuristics operate independently of grammatical competence; standardly, this is represented in the form of a parse-tree (see for example, Marcus 1980) which builds a syntactic structure out of an input string. Of course, this is not all that the parsing component consists of; the Marcus parser, for example, also contains an input buffer which feeds information into the parser and a push-down stack which acts as a type of limited "memory". In other words, in the usual case, the performance component includes the competence component plus more.

All this is pretty obvious, but much that is obvious in less confusing areas of linguistic endeavour loses its clarity in

the twilight world of real-time language acquisition. In particular, there has been a regrettable but widespread and understandable tendency to "save" grammar-based theories of language acquisition by appealing to the influence of unelucidated performance factors. An example is the pro-drop parameter explanation proposed by Weinberg (1987) and discussed in 1.23 above, where a "length-limitation" constraint forces a particular parameter-setting; another, of more venerable vintage (just to show that this style of explanation is not new) is Bellugi's (1968) "sentence programming" explanation for the failure of inversion in children's WH- but not yes/no questions at a certain "stage" in the development of interrogative structures (which we will investigate in detail in Chapter 5). Of course, there is no a priori reason why these accounts should not in fact be correct; our point is merely that without some independent notion of what the performance constraints in question actually look like, it is impossible to assess "explanations" based on them.

In the sub-sections which follow we will attempt to give some substance to the idea of a "performance constraint" by investigating several plausible types of performance-based difficulties in language acquisition; we will take a broad view of the nature of such limitations, including such phenomena as segmentation problems, "pragmatic over-ride", and other "external" factors which would not normally be included under the rubric of "performance".

Initially, we will differentiate four main types of

performance factor: "perception factors", which will include segmentation difficulties and parsing difficulties; "production factors", which will include lexical retrieval problems and "sentence-span" limitations; "complexity factors", which will deal with the issue of semantically-linked performance difficulties; and "interference factors", which will include pragmatic over-ride.

Let us then examine these types in turn.

1.31 Perception Factors

Here, we will deal with two main types of difficulty - those caused by failure to correctly segment an input string, and those caused by failure to parse it properly.

It is generally assumed in constructing a theory of grammatical acquisition that a precondition for syntactic development is the child's ability to analyze connected speech into strings of at least partially labelled words. Less is known about how this is actually accomplished: it has been hypothesized that phonological cues (in particular, those involving prosodic phenomena such as pitch, lengthening and pauses) might signal word boundaries, and even the boundaries of major constituents (see Cooper and Paccia Cooper 1980). Moreover, at least some words appear in isolation in the input, making the child's task considerably easier. However, it can't be that easy, since by now there is a considerable body of evidence showing that children often misanalyze word

boundaries, particularly those involving "cliticizations" of various kinds. Thus K. produced the following questions (and many others like them) around the age of 36 months:

(7)a. And what's the mouse is doing ?

b. What's this is ?

c. What's this can do ?

d. What's these are ?

At this point in his development, K. showed no signs that he had grasped a productive rule of subject auxiliary inversion; indeed, when given yes/no questions to imitate he deinverted them (see Chapter 5 for details). Moreover, these are not instances of so-called "auxiliary overmarking", since K. failed to produce any other instances of "double auxiliary" structures. The only possible (and also, the obvious) conclusion is that what's has not been correctly segmented at this point in K.'s development. This type of segmentation eror is in fact extremely frequent in young children's speech (see Peters 1983 for a number of examples) and, as we shall argue in Chapter 4, can lead to the "mis-setting" of a parameter.

On the other hand, far less is known about whether young children experience parsing difficulties in acquisition: the standard approach (see, for example, Berwick and Weinberg 1984) is to assume that properties of the parser are constant through the course of language acquisition. This is not, of course, a necessary assumption; we could claim, for example, that

children are more limited than adults in their "look ahead" capacity (i.e., the number of as yet unanalyzed units that they can store at any given stage during a parse) or in their "active node stack" capacity (i.e., the number of incomplete units that can be stored at any one time). Either of these modifications would alter rather radically children's ability to parse sentences: essentially, they would be less able to deal with "non-local" dependencies. Thus suppose we were to adopt a parser such as that of Marcus (1980) and limit the capacity of the push-down stack that contains "partially built" constituents to just one unit. In that case, the child would have to "deactivate" (i.e., complete) each unit as it came in. For example, take the sentence dogs chase cats. The child would begin by activating a sentence node S, which would immediately exhaust the capacity of the push-down stack, which can only store one unit. From there on, each unit would have to be parsed immediately, since the parser has run out of storage capacity. In particular, in the example above, the child could not create a VP constituent, since to do so would mean having to put the partially built VP into the push-down stack while analyzing the NP cats. In that case, the child would be forced to create the syntactic representation $[_S$ N V N] rather than the adult structure $[_S$ N $[_{VP}$ V N]]. Even from this simple example, then, it can be seen that altering the properties of the parser can radically affect the syntactic hypotheses available to the first-language learner.

Unfortunately, as far as I know, there has been virtually no

work directed at the exact implications of a "developmentally variable" parsing component. There is obviously a great deal of potential interest here, but all we can do at present is to leave the issue open for future research.

1.32 Production Factors

If little is known about the influence of perception factors on developing grammars, it is fair to say that even less is known about production factors, even though explanations based on "production constraints" are ubiquitous in studies on language acquisition. This subsection will consist almost entirely of hand-waving, with a few pointers for future research; we hope that such research will at least begin to clarify some of the issues raised here.

One of the few "production problems" which have received any attention at all in the language acquisition literature is that of lexical retrieval. It is commonly - and quite reasonably - assumed that certain kinds of lexical information (e.g., subcategorization frames) are activated during both perception and production: this is lexical retrieval. Several models of lexical retrieval have been proposed which account for adult speech errors, most notably that of Stemberger (1982a.), who uses a version of the interactive action model of McLelland and Rumelhart (1981). In this model, semantic information activates the whole of a "semantic field" - i.e., all the words which are positively specified for a given semantic feature; however,

when a complex of features is fed into the system, not all words will be activated equally; those with the most features in common with the input will be activated the most, and their activation will in turn inhibit the activation of other words in the field. Now, when a given word (represented as a bundle of semantic features) is activated, its whole "lexical entry" (i.e., all the syntactic, morphological, and phonological features associated with it) will also be activated; however, all the lexical entries with similar syntactic, morphological, or phonological features will <u>also</u> in turn be activated. Finally (actually, nearly instantaneously), the word whose activation exceeds a given threshold will be chosen.

One of the advantages of this model is that it provides an account of lexical retrieval errors. When the wrong word gets too high a level of activation it will be chosen, either for phonological, semantic, syntactic, or a mixture of reasons, leading to substitution errors; when neither of two words inhibits the other, a mixture of both is possible, leading to so-called blends, or both may be produced, leading to word addition; when both inhibit each other, neither can be chosen, leading to deletions (see Stemberger 1982b. for details). The same approach can of course be extended to other types of lexical information, including subcategorization information.

Quite a number of explanations for error patterns in language acquisition data are based on the supposition that children ocasionally retrieve the "wrong" forms from the

lexicon - "a momentary confusion", to quote Maratsos and Kuczaj (1978); we will be exploring this type of explanation for one set of errors (so-called auxiliary overmarking errors) in Chapter 3. Note that - assuming children's lexicons are structured in a similar way to those of adults - we should expect to find lexical retrieval errors which are typical of adults in children's language as well; though I know of no study which has systematically set out to test this prediction, it seems to be the case that lexical substitution errors at least are reasonably common in children's speech. Thus in K.'s corpus we find examples such as (8):

(8)a. What's this go ? (target: Where's this go ?)
b. What they're going ? (target: Where are they going ?)
c. Where's he pulling ? (target: What's he pulling ?)
d. What they're fighting ? (target: Who are they fighting ?)

These cannot be explained by claiming that K. did not know the meaning of the words what, who and where, since in the vast majority of cases he used all of them appropriately at this stage of development. The likeliest explanation appears to be that they are substitution errors caused by lexical retrieval problems. As for blends, the claim is that auxiliary overmarking structures (which contain two instantiations of tense) are of this type. Turning to the other two (less common) types of adult speech error, I have found no addition errors in K.'s corpus, although it is very difficult to tell what might

count as such, and as for omissions - well, omission is such a pervasive characteristic of young children's speech that to attribute it to "speech errors" is like trying to kill an elephant with a fly-swatter.

To conclude, it appears likely that lexical retrieval errors play a part in the explanation of certain typical error-patterns in language acquisition, although this hypothesis awaits systematic investigation. At least we have a framework in which to investigate the possibility, however, which is more than can be said for the "performance constraint" to which we turn next, which we will call "sentence-span limitation", for want of a better term.

In fact, it turns out that an appeal to sentence-span limitation is perhaps the commmonest - and most traditional - type of performance-based "explanation" in language acquisition; this is paradoxical - or perhaps not so paradoxical - since of all performance constraints, it is the least well-understood. To give a very brief history of its use, it first appeared in the explanation of "telegraphic" speech put forward by Brown and Fraser (1963); to quote Roger Brown, "The adult user of English when he writes a telegram operates under a constraint on length and the child when he first begins to make sentences also operates under a kind of constraint that limits length." (Brown 1973, p.74). This type of account probably reached its most developed state in the work of Bloom (1970), whose method of "rich interpretation" led her to conclude that structures of far greater complexity underlay

young children's speech than ever actually appear - hence the need for "reduction transformations" which obliterate most of the underlying structure to produce typical two- and three-word utterances. Bloom's reason for the discrepancy between underlying complexity and superficial simplicity was, of course, the existence of "performance constraints" - i.e., a sentence-span limitation. Subsequently, Bloom recanted, and child language entered a phase where the prevailing fashion of "semantically-based" grammars for early language made it possible to generate telegraphic speech directly (see in particular Bloom 1973, Bowerman 1973, Schlesinger 1971, Brown 1973). However, during the resurgence of purely syntactic approaches in the late seventies and early eighties (this time married to "semantic bootstrapping", in order to account for the "semantic" appearance of early child language) the sentence-span constraint problem re-emerged, and it is still with us, as witnessed by such very recent proposals as those of Pinker (1984, Chapter 4) and Weinberg (1987), discussed above in 1.23.

Let us now ask the crucial question (since we cannot put it off any longer): what <u>is</u> the sentence-span constraint ? In particular, is it structure dependent (as virtually every other linguistic process we know about is) or does it simply consist of an upper bound to the number of units (words, morphemes, syllables) that a child can produce at any one time ? Is it phonologically sensitive ? Can it be recast as a constraint on "sentence planning" ? Is it reducible to a short term-memory

limitation ? To each of these questions, we must give the same (depressing) answer: nobody knows.

This is not an unavoidable situation. In principle, there are ways to go about answering all of the questions above. Thus we should expect a structure-dependent limitation to be sensitive to certain aspects of heirarchical structure; for example, pieces of constituents should not be eliminated from the surface string. This does not appear to be entirely correct, since the typical "telegraphic" utterance consists of a string of open-class categories (with notable exceptions, such as the "pivot" terms more, other, that, here, there, etc.) minus all inflections, auxiliary verbs, and determiners. Similarly, "phonologically determined telegraphese" should contain all and only phonologically salient elements (of course, this still doesn't answer the question of <u>why</u> utterances should need to be reduced in this way): see Pye (1983) for an approach along these lines. As for the "sentence planning" idea, we need to develop a production model (something like the inverse of a parsing model) which will explain exactly what sentence planning <u>is</u> before we can sensibly discuss such a proposal (though see Pinker and Lebeaux 1981 for some ideas on this subject). Finally, the short-term memory explanation should be testable across cognitive domains, since presumably it is not language specific; moreover, it should apply equally to parsing and production, which would not appear to be the case if children are better at parsing than producing sentences, as is perhaps entailed by the well-known

lag between comprehension and production in first-language acquisition.

Alternatively, we can deny the existence of the sentence-span limitation. This is essentially the position which I will take in Chapter 2, where I will argue that what appears to be the interaction of an extra-linguistic performance constraint and a developing grammar is actually the absence of any grammar at all; to put it bluntly, children at the two word stage have not yet incorporated words into trees, and therefore cannot produce sentences (in the sense that we usually mean by the term "sentence"). This avoids most of the problems presented above, but, as pointed out by Pinker (1984) it does some damage to (a strict version of) the continuity hypothesis, since it implies that there is a "fledgling stage" in language acquisition where the principles of UG have not yet begun to apply. In Chapter 2 we will defend this position, extending Chomsky's (1981) notion of "epistemological priority" from the "logical" to the real-time (one is tempted to say illogical) study of language acquisition, and arguing that structure-dependency is not a property of children's earliest "grammars", which literally do not contain a phrase-structure component.

1.33 Interpretive Complexity Factors

One of the most frequent types of performance factor appealed to in the first-language acquisition literature, and next to sentence-span limitation, probably the least-known, is the idea

of "cumulative complexity", which we have renamed "interpretive complexity".

Just as with sentence-span limitation, this type of explanation has a long history in the child language literature, with its most important exponent being Ursula Bellugi (Bellugi 1968, 1971). Bellugi observed that auxiliaries appeared earlier in young children's speech in yes/no questions than in WH-questions (a generalization we will examine in Chapter 5), and earlier in affirmative than in negative questions. She was led to the conclusion that some measure of "cumulative complexity" might explain this developmental pattern, at first appealing to an acquisitional version of the Derivational Theory of Complexity (DTC). According to this view, two transformations (WH-movement and S(ubject)A(uxiliary)I(nversion)) aree harder to acquire than one (SAI), and three (negative placement, WH-movement, and SAI) harder to acquire than two (WH-movement and SAI). With the collapse of the DTC (see Berwick and Weinberg 1984 for a concise history) this type of explanation fell out of fashion; moreover, since syntactic theory, as usual, had wandered off in a different direction, the simple index of "number of transformations applied" could no longer explain the developmental effect.

However, the child language data did not (and will not) go away; in particular, negation has a strong effect on the probability of various types of syntactic error (including auxiliary overmarking, which will be studied in Chapter 3, and

"non-inversion", which will be investigated in Chapter 5). "Cumulative complexity" explanations are thus left in a kind of limbo between competence and performance; Bellugi (1971) talks vaguely about "limits on the permitted complexity of sentences in the child's grammar" (p.102).

We have no solution for the "complexity" problem; we will remain deliberately neutral as to whether complexity limitations are essentially logico-semantic (for example, it might simply be more difficult to understand the negation of a proposition than the proposition itself) or syntactic (in which case, we might expect them to fall out from some constraint on propositional operators at Logical Form, for example). The label "interpretive complexity" is designed to capture this uncertainty.

1.34 Interference Factors

Under this heading we will include a miscellaneous collection of other possible extra-linguistic influences on children's linguistic development.

Most notable amongst such factors is what we have referred to as "pragmatic over-ride". What we mean by this is the tendency, most often observed in comprehension studies (see in particular Bever 1970 and Maratsos 1974 on the passive and Sheldon 1974 on relative clauses) for "pragmatic plausibility" to over-ride syntactic competence. Given a semantic bootstrapping account of early acquisition (which we will adopt

without justification here; see Chapter 2 for some comments) it is hardly surprising that this shoud occur; there will presumably be a transitional period in the child's development when a mismatch between the semantic/pragmatic reading of a situation and the syntactic/semantic reading of a sentence referring to the situation will lead to the latter rather than the former being rejected: in this case, the child will "trip over the bootstraps".

It is likely that, given a bootstrapping model, "real world plausibility" will influence language development in other ways. For example, the well-known fact that reversible passives are acquired after non-reversible passives (see Maratsos 1978 and references therein) is likely to be linked to the fact that in general the subjects of non-reversible passives are not agents, but themes; this is part of a child's knowledge of the world, not of the language which encodes that distinction. Many other similar examples come to mind, for example, children's early tendency to treat animate arguments as agents and inanimate arguments as themes (see Bowerman 1973).

1.35 Conclusion

In this section we have examined some plausible (and not so plausible) performance factors in language acquisition. The most striking thing about our survey is how little is known about such factors, and how often they are appealed to. Obviously, this is an area crying out for further

investigation. At this point, however, all we can do is minimize our appeal to unexplicated "performance constraints" in explaining data from language acquisition, while remaining aware that such a move is damage-control and not repair.

1.4 On methodology

1.40 Introduction

In this brief section, we will attempt to justify certain methodological assumptions which will be implicit in the rest of the thesis. We do not wish to enter into a full-blown debate about the relative virtues of various naturalistic and experimental methodologies: such debate is beyond the scope of the present study. In the following subsections we will simply describe the most important types of methodology employed in the first language acquisition studies discussed here, with a brief discussion of possible "task-specific" effects, and a few pointers as to what is likely to emerge as the most effective methodological strategy for future work.

1.41 Naturalistic Studies

Nearly 25 years after Chomsky's (1963) dictum urging child language researchers to be "devious", the most important source of data in child language is still the straightforward, single-subject, longtitudinal study of spontaneous production. I would

like to argue that this is in fact as it should be: while there are several disadvantages to naturalistic observation, it seems to me that its role as a yardstick against which to judge other types of methodology is still crucial in the study of children's language. In order to see why, let us first of all list the advantages of this simplest of all methodologies.

First of all, corpora of spontaneous production are easy to collect. This is a purely practical criterion, but important nonetheless, since it means data-collection is not confined to those who can afford sophisticated equipment and research assistants, or to those with special training; this in turn means that there is the potential to build a very broad and uniform data-base. Second, spontaneous corpora are all comparable; there are no problems such as those involved in establishing the equivalence or non-equivalence of data from different experimental methodologies. Third, "experimenter" effects (see Labov 1972) are minimized; spontaneous speech, even in the presence of a tape-recorder, is as close to "natural" as we can get, and an experimenter need not even be present. Fourth, and probably most important, spontaneous speech samples can serve as a "control" for other types of methodology; if a child's ability to imitate, for example, shows no connection to his or her linguistic capacity as revealed in spontaneous speech, we have reason to feel suspicious about the validity of the imitation results as an index of grammatical development.

Let us then turn to the (well-known) disadvantages of

spontaneous speech samples as sources of data for language acquisition. The most serious of these is the same criticism that can be applied to any study based on a corpus of data: there is no guarantee that a (necessarily, finite) corpus will provide a representative picture of the grammar underlying it. This is a particularly acute problem when studying young children, whose spontaneous production is limited in at least the following ways: (a) by the limited range of things they talk about; (b) by the existence of production constraints (see 1.32 above) which prevent them from saying everything they know how to say; (c) by the existence of independent cognitive limitations which further limit their means of expression. Thus it is usually necessary to go beyond the corpus to get a fuller picture of a child's grammar - a move which we will now ourselves make.

1.42 "Enhanced Production" Techniques

The simplest way to "fill in the gaps" left by spontaneous sampling is to attempt to increase the child's production of a given construction type. A range of studies have employed such "elicited production" techniques, including Erreich's (1984) study of the acquisition of questions, which will be examined in detail in Chapter 5. Usually attempts to enhance production involve various human (parents, peers) or non-human (puppets, dolls) interlocutors which are used to manipulate naturalistic speech situations to increase the production of a given

sentence-type. More or less the same advantages and disadvantages as attend "unstructured" collection of spontaneous speech samples apply to these techniques (with the obvious added advantage - if they are successful - of increasing the sample for a given construction-type).

1.43 Imitation Techniques

Turning to more "purely" experimental techniques, let us first examine the methodology generally known as "elicited imitation" (EI), in which, in the simplest case, the child simply repeats a stimulus presented by the experimenter. As an experimental technique, EI has some big advantages: to start with, it is very easy to administer, since even very young children can respond to the simple request "Say what I say". Second, it enables the experimenter to focus in detail on a particular construction type, without having to wait for the child to produce relevant examples; and third, it enables the experimenter to elicit responses for ungrammatical sentences, which are often "normalized" by young children (see for example Kuczaj and Maratsos 1975 and Kuczaj and Brannick 1979, which will be examined in more detail in Chapters 3 and 5, respectively).

However, in spite of its promise, EI has encountered some serious difficulties. The most prominent of these involves the question of whether EI taps the same type of grammatical knowledge as the child exhibits in spontaneous speech; it might

be, for example, that some children are simply good at recalling strings of words, independently of their grammatical capacity. In fact, there is a substantial body of evidence showing that at least in some cases, imitations pass through the child's grammar; this evidence comes from normalizations, where a child will assimilate an ungrammatical sentence to his or her grammatical system, and from the observation that there is a correlation between "imitative competence" and "spontaneous competence" in acquisition, with the former slightly preceding the latter (see Kuczaj and Maratsos 1975).

Nevertheless, these results are only reliable when the length of the utterance to be imitated exceeds the child's capacity to "parrot" - that is, to play back the utterance without running it through the grammatical system. It is known that this capacity varies from child to child and from period to period (see Bloom, Lightbown, and Hood 1975), causing potential problems for cross-sectional EI studies.

There are ways round this problem: it is reasonably easy to establish the child's "parrotting capacity" by pre-testing him or her with strings of unconnected words, and to check the length of utterance to be imitated against a norm established by examining the MLU and UBU (Upper Bound of Utterance) of a given child. In my experience, a child's parrotting capacity seldom exceeds his or her MLU. In addition, children tend to parrot far more often when they don't understand the utterance they are meant to imitate; usually, when I present imitations, I include a context and make sure the child knows all of the

vocabulary items in the string. This also has the effect of reducing the "strangeness" effect which inevitably attends EI sessions.

Note, however, that all of these "checking techniques" depend on the ability to test individual children; for this reason, it appears that EI is a rather more successful experimental technique when applied to small-scale longtitudinal investigations than to large-scale cross-sectional studies. Nevertheless, even with this restriction, it appears to me to be the most promising and illuminating technique currently available; I have made extensive use of it in the investigation of the child K., whose interrogative system is discussed in detail in Chapter 5.

1.44 Comprehension Techniques

While comprehension tests form the backbone of many an experimental acquisition paradigm, they will play a very small part in this thesis. This is not because of any intrinsic weakness, however; rather, it is because comprehension testing depends on an "interpretive mismatch" between stimulus and response, and the types of construction whose acquisition will be discussed here do not typically exhibit such a mismatch. Thus whereas comprehension testing can reveal a great deal about children's interpretation of structures involving interchangeable Õ-roles (e.g., passives) or interpretation of anaphoric dependencies (e.g., pronouns, relative clauses,

control structures) they cannot in principle tell us much about structures which are interpreted correctly but produced incorrectly (such as the "non-inverted" questions we will be studying in Chapter 5).

1.45 Judgement Techniques

As we all know, grammaticality judgements enjoy a privileged place in linguistic theory: indeed, linguistics is unique amongst the cognitive sciences in employing introspection as a serious empirical basis for theory construction. (Would William James have approved ? Who knows ?)

On the other hand, the study of language acquisition is not blessed with such an easily accessible, infallible source of information. It has usually been assumed that young children do not make consistent grammaticality judgements, partially because they do not understand what grammaticality is, and partially because they do not have sufficiently stable grammars on the basis of which to make judgements. Early work on children's metalinguistic abilities (see deVilliers and deVilliers 1974) seemed to confirm this assumption. However, a couple of more recent studies (Kuczaj and Brannick 1979, Stromswold and Pinker 1986) have reintroduced judgement techniques as a viable experimental methodology for tapping the grammatical competence of preschool children (generally, over the age of four), with apparent success. Whether there is a "cut-off point" below which children are unable to produce

consistent metalinguistic judgements, and what the relation of such judgements to other aspects of the children's developing grammatical competence - as revealed by more conventional techniques - might be, are questions whose answers have yet to be determined. Nevertheless, judgement techniques appear to show some promise, at least for the later stages of language acquisition.

1.46 Longtitudinal Versus Cross-Sectional Studies

In this subsection, we will briefly address the issue of the relative virtues of the two main types of sampling employed in first-language acquisition research: many-subject, cross-sectional studies, and small-scale, longtitudinal studies. We will begin with the latter.

The obvious advantage to longtitudinal studies involving the detailed investigation of one or a few subjects is the ability to do far more fine-grained analyses of individual developmental patterns. This applies both "synchronically" and "diachronically"; synchronically, in that the investigator has access to a wide variety of constructions in the child's grammar at any one time - an important consideration, for example, in a parametric model which allows for the possibility of "remote" triggering, where a seemingly unrelated development in one part of the child's grammar can have a crucial effect on another; diachronically, in that the investigator can follow the course of a child's development - and perform

experimental interventions - for as long as is necessary in the study of a particular construction.

In contrast, small-scale studies have one distinct disadvantage; in most cases, they do not allow generalizations across subjects. Thus, in Chapter 5 we will examine a "universal" pattern in the acquisition of questions which was based on the study of three subjects (the Harvard children) and subsequently turned out not to be universal at all.

Turning to cross-sectional studies, we will start with the one big advantage they have over small-scale studies, which is of course, that they allow cross-subject generalizations. Then we will list the numerous disadvantages.

These include the following: depending on experimental design, it is often difficult to discern individual patterns at all, and usually impossible to examine them in any detail - the larger the number of subjects, the greater the difficulty. It is generally impossible to do any follow-up studies with the same children. It is usually difficult to examine more than one construction in any one experiment. It is impossible to establish a rapport with individual subjects, which leads to an intensification of the "strangeness effect" associated with experimental tasks.

Of course, clever experimental design can minimize some of these disadvantages, and analysis of individual response patterns, when presented, significantly reduces the frustration engendered by the attempt to make sense of data which, though subjected to the most stringent of statistical analyses, is

pooled across subjects and therefore virtually uninterpretable. Nevertheless, it is my contention here that a large number of single-subject studies (incorporating experimental techniques, with some standardized procedures for testing and analysis) will always prove more illuminating than a single cross-sectional study using the same number of subjects. There has been a strong psycholinguistic tendency to believe in "safety in numbers" which has led to the publication of many immaculately quantified, theoretically sterile studies; it is to be hoped that as the level of linguistic sophistication of the psycholinguistic community improves, this tendency will be replaced by a much more detailed, fine-grained approach to the investigation of linguistic development. See Peters (1983) for very similar comments.

1.47 Task Specificity and the Cross-Methodological Approach

One of the major problems facing the attempt to probe grammatical competence through experimental techniques is that of task-specificity. It has been known that comprehension and imitation tests tap partially different abilities in the child ever since the pioneering study of Fraser, Bellugi and Brown (1963); indeed, the influence of that study led to a virtual abandonment of the EI technique until its revival by Smith (1970). More recently, Smith and van Kleek (1986) have shown that different types of complexity affect comprehension and imitation tasks: "interpretive complexity" seems relevant to

the former but not the latter.

These results are in fact not very surprising; comprehension tests demand that a linguistic expression be mapped onto a real-world (or at least, make-believe-world) context, whereas imitation tests in general do not. The real question is whether task-specific differences should alter our views about the ability of experimental techniques to probe grammatical competence.

I think the answer to this question is no. As long as we assume that competence underlies both types of performance (since experimental response is, of course, a particular type of linguistic performance), it should be possible to factor out the task-specific elements of comprehension and imitation tests, as long as these remain constant. If there is a moral to task-specificity, it is that no one technique in language acquisition has privileged access to UG (in the way in which grammaticality judgements, for example, have long enjoyed a privileged position in the study of adult grammars). This in turn suggests a "cross-methodological" approach to the study of competence in young children, in which as many experimental techniques as possible are employed, and spontaneous production is used as a kind of yardstick for comparison.

In my pilot study of the developing grammar of K., I have attempted to follow this methodological dictum, introducing various production elicitation, imitation elicitation, and comprehension techniques into free play sessions, sometimes overtly in the form of "language games", sometimes covertly.

Some of the results of this cross-methodological approach are discussed in Chapter 5; a great deal more work needs to be done in order to understand the relationship between the various experimental techniques, but on the whole, the approach seeems promising.

1.48 Conclusion

This has been a very brief and cursory overview of some very important issues. I should like to emphasize that I believe there is in fact a lot more to be extracted from careful analysis of child language data of all kinds - including that obtained through simple samples of spontaneous production - than has often been acknowledged. I thus do not share Chomsky's perennially gloomy view that "...a direct record - an actual corpus - is almost useless, as it stands, for anything but the most superficial linguistic analysis." (Chomsky 1963, p.36). Nevertheless, it is clear that certain outstanding issues - in particular, the delimitation of performance and competence factors in the analysis of data obtained by different experimental techniques - need clarification, if we are to pass beyond the stage of mere observational adequacy towards descriptive and ultimately - somewhere over the rainbow - explanatory adequacy in the study of language acquisition.

1.5 Outline of the Thesis

1.51 Introduction

Having now been briefed on most of the essential concepts and questions likely to be encountered, the reader should be almost ready to embark on the long journey through the rest of this thesis. However, a route map might be very helpful, and accordingly, in this section we will give an outline of the four central chapters.

In fact, the thesis is organized in three parts, each with the same overall structure, beginning with a syntactic description based on English, turning to cross-linguistic analysis, and returning to the acquisition of English. It will be observed that a fourth logical piece is missing from this schema - the analysis of cross-linguistic acquisition patterns. There are two reasons for this. Firstly, including detailed analysis of acquisition data from other languages would have made this thesis unmanageably large. Secondly, in most cases, data of sufficient depth does not yet exist for languages other than English (although this deficit is rapidly being made up). Obviously, such data is extremely relevant to issues addressed here, and I have in fact made use of Clahsen and Smolka's excellent review of the acquisition of INFL2 structures in German and some relevant related work on Dutch (see Clahsen and Smolka 1986) in Chapter 5. A more extended study must await the future.

1.52 Chapter 2

Chapter 2 has two purposes. First of all, it provides an introduction to my own occasionally idiosyncratic version of the theory (or by now, the collection of theories) inspired by Chomsky's Lectures on Government and Binding and related work. Second, it provides an account of the acquisition of phrase-structure which differs rather radically from conventional versions such as that of Pinker (1984). In particular, I argue on the grounds of epistemological priority that phrase-structure is not a primitive component in language acquisition, but is derived from the interaction of other elements of the GB system, in particular Case-theory θ-theory, and a version of "functor theory" (see Abney 1985, Speas 1986). The core concept of government is used as a central organizing relation to map these components onto a phrase-structure representation, NP-structure, via a series of canonical goverment configurations. Crucial to this mapping is an extended version of the Projection Principle of Chomsky (1981). I also introduce a set of percolation principles which determine the "headedness" of phrases; three sets of heads are distinguished: lexical (Õ)-heads, what I call G(grammatical)-heads (a category which includes determiners and inflectional elements) and C-heads, which is a generalized notion of preposition including complementizers (see Emonds 1985).

In the second part of Chapter 2, I compare the theory outlined above with two other contemporary accounts of the acquisition of phrase-structure, those of Pinker (1984) and

O'Grady (1987), before moving onto discuss a particular issue in the acquisition of English, the "missing subject" stage. I argue that a simpler, more accurate description of this phenomenon can be gained by abandoning the "pro-drop" hypothesis of Hyams (1983) and subsequent accounts based on the premise of what I refer to as "fully articulated phrase-structure"; instead, my claim is that the missing-subject phenomenon is presyntactic, in the sense that at the stage in question children have not yet mapped Case and Õ-relations onto a phrase-structure representation.

1.53 Chapter 3

In the first part of Chapter 3, I describe the viccisitudes of the English auxiliary system, arguing that "syntactic" accounts are preferable to "lexical" accounts on the basis of generality both within English and cross-linguistically. I develop a theory of "exceptionality" within the auxiliary system based on the open class/closed class distinction as encoded by Emonds' (1985) Designation Convention, and go on to propose an account of the distribution of auxiliaries and other inflectional elements, based on the theory of categorial association introduced in Chapter 2 and a set of licensing conditions applying to phrase-structure representations at various levels of the grammar.

In the second part of the chapter, I extend this explanation to account for "V-raising" in several languages, including German,

the Scandinavian languages, French, Spanish, and Italian, and develop an explicit parametrization based on this cross-linguistic analysis.

In the third part of the chapter, I return to English to analyze the development of the auxiliary system, arguing that the Designation Convention together with the learning mechanism proposed in Chapter 1 dispenses with the need for explicit "lexical learning" and independently predicts the observed course of development. Finally, I analyze the set of syntactic errors known as "auxiliary overmarking", reviewing several previous accounts before proposing my own, based on an interaction of performance and competence factors.

1.54 Chapter 4

In Chapter 4, I take on the task of accounting for Subject-auxiliary Inversion (SAI) and related rules, including the so-called V2 condition in Germanic languages, Complex Inversion and Stylistic Inversion in French, and "V-preposing" in Spanish and other Romance languages. Since no satisfactory unified account currently exists within GB theory, I invent one, based on the notions of "P(redication)-subject" and "AGR(eement) domain"; these concepts are related to the notion of "I-subject" independently proposed by Borer (1986).

The first part of the chapter deals with SAI in English, and as in Chapter 3, the second part extends the account to a range of European languages, with the aim of providing an

explicit parametrization of inversion principles.

1.55 Chapter 5

In Chapter 5 I return to language acquisition once more, exploring in some detail the acquisition of SAI in English. On the basis of a wide range of published and unpublished studies, I dispute the claim, first made by Bellugi (1971) that the acquisition of SAI involves a series of invariant "stages"; instead, I propose an alternative based on a particular "misprojection" of the PLD available to children learning English. I also compare the acquisition of SAI with data from the acquisition of INFL2 in German and Dutch, and show that the differences can be accounted for on the basis of the view of "acquisitional markedness" proposed in Chapter 1. Finally, I propose a "two-tiered" model of first-language acquisition, designed to overcome some of the limitations of the linear-parametric model without at the same time losing the generalizations it was originally designed to capture.

1.56 Conclusion

What will all this accomplish ? Well, there are at least three different sets of substantive hypotheses embodied in this thesis:

(a) There are a set of purely syntactic analyses: of phrase-structure, of auxiliary systems, and of SAI-type rules.

(b) There are a set of parametrizations, based on extensions of these syntactic analyses.

(c) There are a set of analyses of language acquisition data, based on a complex interaction of language-particular, universal, and language-independent factors.

The idea is - and it is an ambitious one - that on the basis of (a.- c.) it should be possible to provide at least tentative answers to a set of higher-order questions, concerning the relation between the parameter-setting model of language acquisition, cross-linguistic generalizations, and real-time language acquisition, which have been briefly discussed in this chapter. Of course, any of (a.- c.) - actually, probably all of them - could be wrong. This is a risk we take; the enterprise itself, we are convinced, is worthwhile, either as a success or as a failure.

CHAPTER 2

GB THEORY AS A THEORY OF LANGUAGE ACQUISITION

2.0 Introduction

In the last chapter we introduced and discussed a parameter-setting theory of language acquisition. In this chapter we will examine a particular grammatical framework - the Government and Binding (GB) model of syntax - from the perspective of language acquisition. We will be particularly concerned with "epistemological priority" in the sense of Chomsky (1981,p.10):

> "...we want the primitives to be concepts that can plausibly be presumed to provide a preliminary, prelinguistic analysis of a reasonable selection of presented data, that is, to provide the primary linguistic data that are mapped by the language faculty to a grammar..."

Under this conception, principles of UG are principles of language acquisition which enable the child to map pre-linguistic representational systems onto the grammars of individual languages. In order to do this, these principles must be able to match elements and structures from prelinguistic systems with elements and structures in linguistic systems. Principles which have this ability have epistemological priority over principles which are defined in purely linguistic or purely non-linguistic terms. One of the main goals of this chapter is to present a version of GB theory

which is epistemologically plausible; this will involve rather radical changes in some of the grammatical subcomponents of the system. In particular, we will attempt to derive the principles of X-bar theory from the interaction of more primitive components of the system, most notably Õ-theory, Case-theory, and a version of "functor theory" (see Abney 1985, Speas 1986). This will involve a new syntactic feature system and a set of categorial feature-percolation principles, and will have widespread effects on other subcomponents of the GB system.

The structure of the chapter will be as follows: in sections 2.1 and 2.2 the overall architecture of the model will be reviewed and the fundamental structural concept of government introduced; in sections 2.3 - 2.8 the structure of the various subcomponents will be introduced and discussed; and section 2.9 will be devoted to an examination of learnability considerations and the relation of the model proposed here to some aspects of real-time language acquisition.

2.1 Grammatical subcomponents

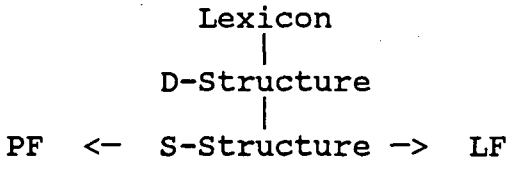
2.10 Architecture of the GB model

The GB theory of syntax is "modular" in that several distinct subcomponents interact to determine the well-formedness of a given configuration. These subcomponents include

(1) X-bar theory.
θ-theory.
Case-theory.
Bounding Theory.
Binding theory.
Government theory.

Each of these subcomponents contains principles and associated parameters. The grammar is organized so that each subcomponent has a different locus of application within the "Y-shaped" model adopted by most GB theorists:

(2)



(PF="Phonetic Form", LF="Logical Form")

A given principle may apply at any or all of the levels distinguished above, though intuitively there is a certain division of labour between those which concern the phonetic realization and the semantic realization of a given grammatical element.

We will be modifying this model slightly during the course of this chapter, arguing in particular for a level of "NP-structure" intermediate between D-structure and S-structure, as first proposed by van Riemsdijk and Williams (1981).

2.11 The Projection Principle

An important metatheoretical principle, the Projection

Principle, constrains the overall structure of the grammar. In its "extended" form (see Chomsky 1982, Chapter 1), the Projection Principle (PP) has two parts: firstly, it requires that lexical (i.e. subcategorization) requirements be satisfied at each level of the grammar - except, obviously, at PF - (thus a syntactic structure is said to be "projected" from the lexicon); secondly, it requires that all clauses have a subject at each relevant level. Together these two conditions allow the child (and the linguist) to infer the existence of phonologically null elements (empty categories) in positions where the PP requires that an element be present.

We will be proposing a revised and extended version of the Projection Principle in 2.56, which will include subcategorizing as well as subcategorized elements. This principle, the Generalized Revised Extended Projection Principle (GREPP) will play an important part in distinguishing what we shall call the "subcategorization set" and the "adjunct set" of a phrase; this distinction in turn is crucial in our model for the learning of phrase-structure.

The PP has had a very controversial history. Since it is so closely tied to questions of learnability, we will postpone a more detailed discussion of its implications until after we have presented the other subcomponents of the GB system, at which point we will be in a better position to assess its contribution to the overall theory.

2.2 Government

2.20 Overview

Before turning to the subdomains themselves, it will be necessary to give a brief introduction to the basic structural relation of government. In most versions of GB, government could be said to impose a general condition on the mapping of the various syntactic subcomponents onto each other: thus θ-roles and abstract Case are usually assigned under government; the notion of governing category is crucial to binding theory; several versions of bounding theory (e.g. those of Kayne 1984 and Chomsky 1986b.) make use of government; and the central principle of "Government Theory", the Empty Category Principle, depends on a restricted version of government known as proper government. However, in spite of (or perhaps because of) its importance, government remains problematic; though it is a potentially unifying concept, no unified version of the concept itself seems to be forthcoming. One of the major aims of this chapter is to show that the idea of a unified concept of government is in fact mistaken, and that different types of government relation characterize different syntactic components. In section 2.9 we will argue that such a relativized version of government can lead to interesting predictions about the course of real-time language acquisition.

Firstly, for expository reasons, let us distinguish two forms of government, which we will call internal and external government. Internal government concerns the relation holding

between a lexical governor and elements within its maximal projection; external government concerns the relation holding between a governor and elements within a maximal projection which it governs. Let us deal with these two cases one at a time, beginning with the simpler case of internal government.

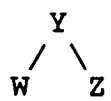
2.21 Internal Government

In the core case, a lexical head governs its complements. Thus in (3) below the verb governs its direct object and the preposition its prepositional object:

(3) I gave a bone to the dog.

Notice that in the core case, θ-role assignment and Case-assignment coincide, but that in non-core cases they need not; thus the subject NP in the sentence above receives its θ-role from the VP (possibly via VP government - see Travis 1984) but its Case from INFL. We see that the core cases involve the configuration in (4), where W is a θ-assigning, Case-assigning head, Z is its complement, and Y is a projection of W:

(4)



Adapting an idea first proposed by Kayne (1984), let us say that this core case represents a Canonical Government

Configuration (CGC), as defined below in (5):

(5) W and Z, immediately dominated by some Y, are in a Canonical Government Configuration iff

a. V Case- and θ-marks NP to its right in the grammar of the language in question and W precedes Z or

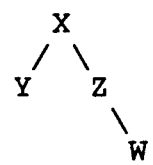
b. V Case- and θ- marks NP to its left in the grammar of the language in question and Z precedes W.

We will make crucial use of a much-modified and relativized notion of CGC in our account of the acquisition of phrase-structure, where government will be used to map Case- and θ-relations into X-bar configurations.

We will refer to this core type of government as minimal government, as opposed to maximal government, to which we now turn.

Let us examine the configuration in (6):

(6)



Here, W is a lexical governor, Z and X are projections of W, and Y is a "specifier". The question is, does W govern Y ?

We have seen that in the core case of the CGC, the environments for Case- and θ-role assignment coincide. Let us then examine the Case- and θ-relations holding between W and Y in (6), with a view to determining whether these cases can be included under a straightforward extension of the CGC.

There are two types of case to consider; in the first, Y is a subject, and in the second, Y is what we will provisionally term a "non-subject specifier". Let us deal with these one by one.

First, let us take Y to be a subject. Again, there are two main cases to consider: Y may be the subject of S, or the subject of NP. Suppose Y is the subject of S. Then under current GB assumptions, W = INFL, Z is an intermediate projection of INFL, and X (= S) is the maximal projection of INFL. Now, the subject in Y does not receive a θ-role from INFL, but from V (or more likely compositionally from VP; see Marantz 1984). Thus in this case a θ-role cannot be assigned under minimal government. On the other hand, if we accept the common view that nominative Case is assigned by INFL, and continue to assume that Case is uniquely assigned under government from a lexical head, then we must conclude that INFL governs the subject after all. If so, however, note that this is a non-canonical type of government, since INFL assigns nominative Case to the left, whereas in the CGC cases, government is rightward.

Turning to NP, if W is N, then Y may again be a subject (this time of NP) and again, it will be in a Case- and θ-marking environment. We can assimilate the assignment of the "external" θ-role in NP to that of the external θ-role in S by assuming that N' optionally assigns a θ-role to the subject of NP under predication (though see Williams 1982 who argues that there is only a vague "aboutness" connection between the two,

and Rothstein 1983 who argues that predication does not take place within NP). However, when we turn to Case-assignment in NP, there is some controversy over whether genitive Case is assigned under government at all; Chomsky (1981), for example, claims that genitive is assigned "structurally", mainly on the basis of NP gerunds such as (7), where there appears to be no nominal head to assign genitive Case:

(7) John's singing the Missa Solemnis in the bath drove everyone to tears.

We will not adopt this view, but instead assume that genitive Case is one realization of the functional category DET(erminer), and is assigned, like nominative Case, in a non-canonical government configuration (see Abney 1985, Stowell 1987).

Comparing government of the subjects of NP and S, we note that in both NP and S, Case- and θ-role assignment is uniformly to the left, whereas in the CGC cases, both are to the right. This suggests that if a government relation does hold between heads and subjects, it is of a different type than the minimal Government relation holding between heads and complements. Let us accordingly name this type of government <u>maximal government</u>. In fact, maximal government corresponds to the widely accepted version of government first proposed by Aoun and Sportiche (1983), where a governor governs everything up to a maximal projection. In subsequent sections (2.3 and 2.4, in particular)

we will argue that there is a systematic correspondance between what is generally known as "internal" Case and θ-role assignment (not to be confused with our purely heuristic distinction between internal and external government) and minimal government, and between "external" Case and θ-role assignment and maximal government. Furthermore, in section 2.9, we will show that this distinction is important to the acquisition of phrase-structure, where government relations effectuate a mapping between Case- and θ-relations and phrase-structure configurations.

Having examined the relation between subjects and lexical heads, let us then briefly examine the relation between "non-subject specifiers" and heads, a topic to which we will return in more detail in 2.5.

Non-subject specifiers include determiners and quantifiers (in NP), auxiliary aspectual verbs (in VP), intensifiers and measure phrases (in AP), and, in many current versions of GB including that of Chomsky (1986b.), WH-phrases (in (SPEC) COMP). Now, it is clear that none of these elements receive (primary) θ-roles; there is, however, evidence that some of them may receive Case in exceptional circumstances (see 2.3 below). Interestingly, in these exceptional circumstances, the Case-assigner is invariably a lexical governor and Case-assignment is invariably to the right; i.e., exceptional Case-marking takes place in a CGC, under minimal government. Non-subject specifiers never receive Case from "below"; i.e., in configuration (6), W never assigns Case to Y if Y is a non-

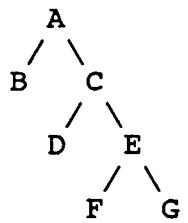
subject specifier. It would thus seem that there is little to be gained from treating W as a governor for a non-subject specifier Y in (6), since there is no evidence of a Case- or θ-relation between the two.

Nevertheless, there do seem to be close relations between instantiations of W and non-subject Y in (6). In section 2.5 we will propose a radical reformulation of X-bar theory to capture these relations, doing away with the problematic notion of "specifier" altogether and relativizing the notion of "head" to include non-subject specifiers.

2.22 External Government

We now turn to external government. As aptly described by Chomsky (1986b.), this is a murky area, into which we will attempt not to sink too far. The relevant configuration is as in (8) below:

(8)



Here B is a potential governor in a CGC with its "complement" C, D is a "specifier", and C and E are projections of the lexical head F. The question is: given that B governs C, which of (D,E,F,G) does it govern ?

The answer to this question in part depends on the "permeability" of C; and this in turn depends on whether C is lexically selected ("L-marked") and whether C can constitute a "barrier" to government (S, for example, does not count as a barrier in the theory of Chomsky 1986b.). Leaving aside for the moment these difficult questions (see Chomsky, op. cit., for details, as well as sections 2.3, 2.7, and 2.8 of this chapter), let us assume that C in (8) above does not constitute a barrier. What then does B govern ?

There seems to be good evidence that it governs D (the "specifier" of F). Suppose we take B to be an exceptional Case-marking verb and C to be S (as in Massam 1985). Then, if we assume that exceptional Case-marking takes place in a CGC (see 2.4), V governs SPEC I, the embedded subject, to which it assigns objective Case.

There also appears to be evidence for Case-assignment to a WH-phrase in S'; the following example (from Kayne 1984) demonstrates that it is possible for a WH-word originating in an environment of nominative Case-assignment to surface with objective Case:

(9) Whom$_i$ did you believe [e_i [the police suspected [e_i [e_i was the criminal]]]] ?

Apparently the only way the WH-word can get objective Case is via one of the intermediate traces in "COMP", which must then be Case-accessible to a governing V. If we assume that such

traces are in a "SPEC COMP" position (see Koopman 1984 and Chomsky 1986b.) then we have further evidence for external Case-assignment to a specifier position.

It is less easy to show that Case is assigned externally to the "specifier" of N; however, note that overt objective Case shows up only on pronouns, which have been analyzed by Postal (1966) as occupying the SPEC N position rather than the head N position, on the basis of sentences such as (10) below:

(10) You two go on ahead, and we three will wait here.

If the pronouns in such cases are in a specifier position, then examples such as (10) support the notion of external government to D in (8).

Another source of evidence for external government of D in (8) is lexical selection. If we assume that syntactic features are assigned under government, and that "external" government always involves minimal government, then the fact that they frequently turn up on D should come as no surprise. A few examples: selection of the feature [+WH] by a governing V can be satisfied either by the appearance of a WH-phrase in SPEC C or by a WH-complementizer in the COMP position; selection of a [+DEF] NP can be achieved either by the use of a definite article in SPEC N or by the use of a proper noun in the head N position; tense features can appear either on aspectuals in the SPEC V position or on the head V itself.

We thus conclude that there is overwhelming evidence for

allowing external government of the specifier-position D.

Next we turn to government of F by B in (8), where F is itself a governor in a CGC with G. This is the core case of "percolation government" (Kayne 1984); it is generally assumed (following Belletti and Rizzi 1981) that θ-roles are assigned under government to a complement XP (i.e., C in (8) and percolate (via the intermediate projection E) to the "thematic head" F of that phrase, in which case "head-to-head" government must be admitted. We conclude that F may be externally governed by B in (8).

Conversely, there is little controversy over whether the complement position G should be externally governed; it is unanimously assumed that it should not. In Chomsky (1986b.) this is achieved by a "minimality condition"; the basic idea is that "government domains" cannot overlap, (see also Reuland 1983) so that a lexical governor blocks government by a more "distant" governor.

We now have a characterization of both internal and external government. We have distinguished two types of government, minimal and maximal, which may be captured by the following definitions:

(11)a. Minimal government:

> α minimally governs β iff α minimally c-commands β and there is no Γ such that α governs Γ and Γ governs β.

b. Maximal government:

α maximally governs β iff α maximally c-commands β and there is no Γ such that α governs Γ and Γ governs β.

Further variation in government involves the specification of Γ; in other words, different elements create "opaque government domains" for different types of government, as we shall see.

With this (by no means unproblematic) working version of two types of government, let us go on to explore some of the subcomponents of the GB system.

2.3 θ-Theory

2.30 The θ-Criterion

At the heart of the GB approach to the assignment of θ(thematic)-roles is a biuniqueness condition, the θ-criterion. As stated in Chomsky (1981, Chapter 2), the θ-criterion holds that

(12) Each argument bears one and only one θ-role, and each θ-role is assigned to one and only one argument.

(Note however, the more complex reformulation in Chomsky 1981, Chapter 6, which refers to argument chains rather than argument positions, and thus avoids many of the problems

alluded to below. We will retain the simpler version for reasons of expository clarity.)

This condition has the effect of preventing movement from one θ-marked position to another, since movement creates a chain of coindexed positions (traces) and the θ-criterion will rule out any chain which bears more than one θ-role. Because complement positions always get assigned a θ-role, movement into such positions will always violate the θ-criterion; thus, a raising-to-object analysis will be impossible for (13) below, since expect would then assign a θ-role to its direct object, which has already been assigned a a θ-role by go berserk in the embedded clause:

(13) I expect him to go berserk any minute now.

Instead, an operation of "exceptional Case-marking" is employed, whereby the matrix verb assigns objective Case to the embedded subject, which it governs by (12a.).

Raising-to-subject, on the other hand, is permitted as long as the subject in question has not been assigned a θ-role. Monadic predicates such as seem do not assign a θ-role to their subject (which has to be present whether θ-marked or not, since it is required by the Extended Projection Principle); the subject position is thus free as a landing site for NP-movement, as in (14a.), which in most GB accounts is derived from a D-structure such as that in (14b.):

(14)a. He seems to have published many articles.

b. [e seems [he to have published many articles]]

A similar analysis is provided for passivized structures, where the θ-role assigned to subject position is supressed, and to "unaccusative" verbs, which (following Burzio 1981) are analyzed as monadic predicates which assign a θ-role only to an NP in complement position.

In terms of learnability, it has been claimed that the θ-criterion helps to restrict the language learner's hypothesis space by restricting movement to non-θ (θ') positions.

θ-theory also plays a possible role in limiting phrase-structure configurations; recent work by Travis (1984) argues that θ-role assignment is directional, and that the interaction of directionality parameters for Case (see below) and θ-role assignment fully determines word order.

Perhaps unfortunately for such arguments, there is evidence that Chomsky's "strong" version of the θ-criterion is <u>too</u> restrictive (again, we note that the relativized version in Chomsky 1981, Chapter 6, and Chomsky 1986a. avoids the problems we are about to discuss; however, since we are abandoning PRO in any case, this is largely irrelevant to our current concerns). This evidence comes from "defective predicates" which do not appear to function as clauses at the syntactic level, though they do get interpreted as such at some semantic level. These predicates are of two types. The first are adjunct

predicates (so-called "small clauses") of the type exemplified below:

(15) John ate the meat raw/nude/in a hurry.

Here, the θ-criterion can only be satisfied if an empty subject pronoun (PRO) is postulated for the θ-role assigning adjunct predicate phrase; otherwise, the matrix subject will receive two θ-roles, in violation of the second part of the θ-criterion. Thus in the "small clause" analysis advocated by Stowell (1983), sentences such as (15) will be analyzed as in (16) below:

(16) John [[ate the meat] [PRO raw/nude/in a hurry]]

However, Williams (1980, 1983) provides several arguments against this type of structure. While it is beyond the scope of this thesis to enter into a detailed discussion of this issue, the basic thrust of the argumentation is to show that adjunct predicates behave differently in the syntax from "true" clauses, and treating the two as structurally parallel merely obscures these differences.

The second type of defective predicate is exemplified by certain types of non-clausal complement. These include non-nominal gerunds, complements to verbs of perception and temporal aspect, and the complements to <u>make</u> and <u>let</u>. These types of complement differ syntactically from full clausal

complements in several ways: see Emonds (1985, chapter 2) for a thorough exposition.

Now, if Williams and Emonds are right, then we must admit the possibility of subjects receiving θ-roles from more than one predicate. Since the strong version of the θ-criterion is a bi-uniqueness condition, it may be broken down into its constituent parts:

(17)a. Each argument bears one and only one θ-role.

b. Each θ-role is assigned to one and only one argument.

It is (17a.) that must be modified. Various proposals have been put foward as to how to reformulate the θ-criterion without it: see Williams (1983), Emonds (1985), and Chomsky (1986a.) We will simply drop (17a.), adopting a "weak" θ-criterion consisting of (17b.) alone.

This relaxed version of the θ-criterion permits a non-clausal analysis of infinitival complements, since PRO is no longer necessary to receive the subject θ-role of an infinitival verb. In fact a non-clausal view of infinitivals is espoused by many of the more "concrete" grammatical theorists, including practitioners of Lexical Functional Grammar and Generalized Phrase Structure Grammar, and has been adopted in some versions of the Extended Standard Theory, notably that of Culicover & Wilkins (1984). We will propose a variant of this approach, in effect claiming that infinitives are headed by an instantiation of INFL (to) which does not license a subject.

Since Emonds (1985) explicitly argues for a clausal analysis of infinitives, we will address his arguments here.

2.31 On PRO

Following Koster and May (1982), Emonds presents eight arguments relating to the "Understood Subject Property" (USP) of infinitives, purportedly showing that an empty subject must be present in infinitives but not in non-NP gerunds or bare VP complements. Let us review these arguments one by one.

(i) Infinitives may have lexical subjects which receive a θ-role from an embedded predicate alone, as in Exceptional Case-marking environments:

(18) John considers Bill to be a maniac.

Here Bill is not assigned a θ-role by the matrix verb but only by the embedded predicate. This argument also applies to sentences such as (19), however:

(19) John considers Bill a maniac.

Exactly the same θ-relations hold in (19) as in (18), though according to Emonds' own hypothesis, the complement "phrase" in (19) is not clausal, but consists of two NPs. In other words, the "lexical subject property" cannot be used as a

diagnostic for clausal status. Furthermore, Emonds' claim that non-NP gerunds never have independent lexical subjects is falsified by absolutive gerunds, which do:

(20) John having lost his keys, none of us could get into the room.

(ii) Following Koster & May, Emonds claims that to is an auxiliary, since it controls an empty VP in "VP-deletion" environments:

(21) John wanted to scream, and Bill wanted to as well.

However, while we accept that to is an auxiliary, it does not follow that it must license a subject.

(iii) Infinitives, like tensed complement clauses but unlike (non-NP) gerunds, may appear as the focus of pseudo-clefts:

(22)a. What John imagined was that monsters were pursuing him.
b. What John decided was to hide under the bed.
c. *What John continued was sweating like a pig.

However, any analysis which treats to as an inflectional element correctly predicts the pseudocleft facts, since non-NP gerunds are VPs, which do not appear as the focus of pseudo-clefts, whereas infinitivals are a special type of

I(nflectional) P(hrase). It might be argued that only S' (i.e., C(omplementizer) P(hrases) are possible in the pseudocleft position (that is obligatory in (22a.); here, following Chomsky (1981) and Henry (1987), we assume a for-deletion rule which converts the infinitival CP for to hide under the bed in an example such as (22b.) into the "bare IP" to hide under the bed at PF.

(iv) Infinitives may be conjoined with tensed S's (C"s) but gerunds may not be coordinated with C"s or infinitives:

(23)a. John believed his doctors to be vampires and that they would take all his blood.

b. *John tried screaming and to call his mother.

c. *John considered running away and that noone would catch him if he ran fast enough.

However, under our analysis, a to phrase is a reduced CP; thus the facts fall out without further stipulation.

(v) Infinitives, like tensed clauses but unlike gerunds, may appear with an overt preposed WH-phrase:

(24)a. I saw what to do.

b. I saw what John had done.

c. *I saw what (John) doing.

Again, this array of data can be accounted for by assuming that

a to-phrase is a reduced CP, which licenses the presence of a WH-phrase; there is no need to invoke an understood subject.

(vi) Certain anaphoric expressions (reflexives and reciprocals) normally require a c-commanding antecedent. However, they turn up in infinitivals without one:

(25)a. To hurt each other would upset Mary and Bill

b. To see himself now would make Bill cringe.

This otherwise surprising data is easily accounted for if we assume the presence of PRO. However, it is not the case that there is no antecedent present in the examples in (25); these are both cases of "obligatory control", as evidenced by the examples in (26):

(26)a. *To hurt ourselves would upset Mary and Bill.

b. *To see himself now would make Bill's girlfriend cringe.

Since obligatory control requires a c-commanding antecedent (see Williams 1980), the problem of "antecedentless anaphors" cannot be solved by merely assuming a PRO subject in the infinitival subjects in (25); this merely transfers the problem to a (largely promissory) theory of control. Moreover, since non-NP gerunds never show up as "sentential" subjects (not surprisingly; only NPs and marginally PPs may occupy subject position) sentences with antecedentless anaphors cannot

distinguish gerunds from infinitivals. Of course, NP-gerunds freely occupy subject position, and when they do, the problem resurfaces, as in (27):

(27)a. Seeing myself in the mirror makes me vaguely suicidal.

b. *Seeing myself in the mirror is enough to make anyone sick.

Now, in order to handle these cases in Koster & May's terms, a PRO must be present as the subject of NP; yet in general, the subject of an NP is optional, even when the head N assigns an external θ-role (see Williams 1982). Note that an extension of this analysis requires a PRO in sentences such as (28):

(28)a. All those books about themselves just inflate the egos of politicians.

b. Hatred of each other prevents many academics from exchanging ideas.

Note that the only motivation for postulating PRO in these environments is the appearance of antecedentless anaphors, and that the structure of (28) will include a determiner as well as a subject, a configuration normally ungrammatical in NPs:

(29)a. *Those Dostoevsky's novels move me to tears.

b. *Dostoevsky's those novels move me to tears.

We thus conclude that however antecedentless anaphors are to be treated, postulating a controlling PRO does not contribute to a

solution nor provide any arguments in favour of PRO in infinitivals.

(vii) Following Koster & May, Emonds claims that a PRO allows each other to have a split antecedent in infinitives but not in gerunds, as illustrated in (30) below:

(30)a. ?John proposed to Mary to help each other.

b. *John suggested to Mary helping each other.

I find both these examples unacceptable; certainly, the following (exactly parallel) cases are ungrammatical:

(31)a. *John asked/told/persuaded Mary to help each other.

b. *John argued/pleaded/agreed with Mary to help each other.

Moreover, while Emonds claims that non-NP gerunds never allow split antecedent constructions, the following examples seem entirely comparable in (un)acceptability to the infinitival cases:

(32)a. ?*The truck collided with the car, destroying each other.

b. ?*The Russians met with the Americans, helping each other to avert nuclear disaster.

There thus appears to be no basis for deciding between gerunds and infinitives as far as split antecedents are concerned.

(viii) Finally, as Koster & May point out, infinitives and tensed clausal complements have essentially the same distribution. This is accounted for straightforwardly in the system we will be adopting, since both INFL and infinitival to share the category feature [I], which gerunds do not have; and both tensed and untensed complement clauses are instantiations of CP, though the for deletion rule (which is subject to dialectal variation: see Henry 1987) disguises this fact in infinitivals.

We conclude that there are no strong structure-based arguments for the existence of PRO. Since we have weakened the θ-criterion to allow non-clausal V, CP and IP complements, theta-theory provides no conceptual grounds for retaining PRO either.

One last potential problem remains, however: this is the phenomenon of arbitrary control. As is well known, under certain circumstances the understood subject of a VP need not be present at all in the syntax; this is illustrated in (33) below:

(33)a. To be, or not to be, that is the question.

b. Writing a thesis can be the most frustrating of occupations.

c. The threat to use nuclear armaments should not be taken lightly.

d. What to do about it, however, is a difficult question.

In order to account for such examples, I'd like to turn the problem on its head, and claim that no VP obligatorily assigns a (syntactically realized) θ-role. In other words, in this respect, S and NP are parallel; as can be seen from (34), N' need not assign a θ-role either:

(34)a. The destruction of cities is an appalling crime.
b. Destroying cities is an appalling crime.
c. To destroy cities is an appalling crime.

Of course, we now have the problem of accounting for why in the normal case, VP does appear to assign an obligatory θ-role. This, I claim, is due to a quite separate set of conditions governing syntactic predication, which we will present in 2.33, and which crucially depend on the presence of AGR(eement) features.

2.32 θ-Government

While the weak version of the θ-criterion allows us to account for various types of non-clausal complementation, it also allows many other analyses that the strong θ-criterion would have prohibited. For example, there is now nothing in the θ-criterion to prevent raising-to-object analyses, since nothing prevents an NP from receiving more than one θ-role (such structures will however be independently ruled out by the Projection Principle). It is not at all clear that this is

necessarily a disadvantage; if anything, it eliminates a redundancy in the system. In fact, in order to analyze the complement structure of verbs like ask or persuade without employing PRO, we will have to postulate an object NP which receives two θ-roles, one from the matrix predicate, one from the embedded predicate. This is illustrated below:

(35)a. I asked him to go to bed immediately.

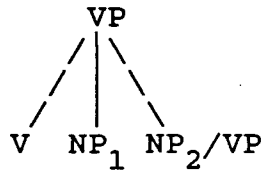
b. I persuaded him to go to bed on time.

c. I asked him the time.

d. I persuaded him of the importance of a good night's sleep.

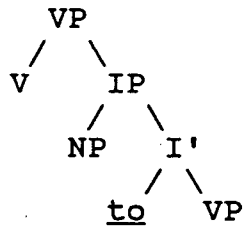
As we can see from (35c.) and (d.), both ask and persuade assign two complement θ-roles. Since these are assigned under government in a CGC, we come up with the following VP structure:

(36)



When a persuade-type verb assigns a θ-role to a VP, this θ-marked VP must also assign a subject θ-role to the complement NP_1, which consequently receives two θ-roles, violating the "strong" θ-criterion. Now compare this type of complement-structure to that of believe or expect:

(37)



Since believe-type verbs only assign one complement θ-role, and since complement θ-roles are assigned in a CGC, IP will be θ-marked, but NP will not be. Of course, this is exactly the result we wish to obtain; the matrix verb in such constructions imposes no selectional restrictions on the embedded "subject", which receives its single θ-role from the embedded VP.

The structures we have proposed in (36) and (37) capture the well-known differences between persuade-type verbs and believe-type verbs, for example, their behaviour when passivized:

(38)a. I persuaded the doctor to examine Bill.

b. I believed the doctor to have examined Bill.

c. I persuaded Bill to be examined by the doctor.

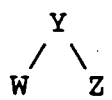
d. I believed Bill to have been examined by the doctor.

As shown by (38), when the complement of a believe-type verb undergoes passivization, the resulting structure has an identical interpretation to its non-passivized counterpart, whereas passivized complement structures in persuade-type verbs have a different interpretation from their non-passivized counterparts. This is easily explicable in terms of the analysis we have proposed: both the "original" subject and the

"derived" subject of the persuade-type complement also receive the goal θ-role assigned by the matrix verb, whereas neither the original object nor the oblique agent of a passivized structure receives any extra θ-role.

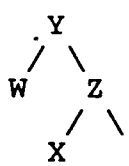
Though we have appealed to a structural difference in accounting for the divergent behaviour of persuade- and believe- type verbs, this difference is not basic to our account, but is itself derived from the different θ-role assigning possibilities of the two types of predicate, and more particularly, the structural instantiation of such possibilities. Thus we have made crucial use of the fact that θ-role assignment in a CGC invariably occurs in the minimal government configuration

(39)



Now, notice that this configuration is even more restricted than the one we proposed for Case-assignment, where, as in (40) below, the "specifier" X is accessible to a Case-assigner, as for example in Exceptional Case-marking structures:

(40)



In other words, there is no "Exceptional θ-role assignment". Let us then distinguish "θ-government", which takes place in

(39), from "Case-government", which takes place in (39) and (40), noting that the former is a subset of the latter; we shall make use of this subset relation in our account of the acquisition of phrase-structure.

2.33 Predication

Most accounts of predication are "θ-based" (see for example Williams 1980 and Rothstein 1983); as such, they are primarily concerned with θ-role assignment to "external arguments"; i.e., subjects. In this section we will introduce a quite different, "autonomous" theory of predication, including both "A-type" and "A'-type" predication; this will provide the basis for our account of certain types of inversion rule in Chapter 4. As for external arguments, the claim made here is that the theory of predication is only <u>indirectly</u> connected to θ-role assignment, and thus that much of what has hitherto been accounted for in predication theory is actually non-syntactic in nature.

We begin by presenting the three central "licensing conditions" of the theory to be elaborated:

(41) I. Every predicate must be contained within an <u>AGR-domain</u>

II. Every <u>AGR-domain</u> must contain a <u>Predication Subject</u>.

III. Every predication subject must be <u>AGR-linked</u>.

The three conditions in (41) obviously depend on prior definitions of AGR-domain, AGR-linking, and P(redication)-

subject. Let us assume for the moment that a P-subject is an NP assigned nominative Case by INFL (the normal case of an A-type P-subject). AGR-linking will be defined as below:

(42) AGR-linking: a P-subject α is linked to AGR_i iff α minimally governs AGR_i.

The definition of minimal government assumed here is that of (11a.); this entails the minimality condition discussed in 2.2, so that a P-subject will create an "opaque government domain" for another P-subject. AGR is assumed to be a feature complex contained in INFL, a reasonably standard assumption (see Chomsky 1981).

Next, we turn to the notion of AGR-domain:

(43) α is in the AGR-domain of AGR_i iff α is in the maximal-government domain of AGR_i.

Once again, the maximal government definition of (11b.), including a minimality condition, ensures that AGR-domains do not overlap. In Chapter 4, when we deal with A'-type predication, we will propose a more elaborate version of (43), incorporating Chomsky's (1986b.) notion of "barrier"; however, for present purposes, the simpler version proposed here is adequate. Note that predication involves both minimal government (in AGR-linking) and maximal government (in the definition of AGR-domain).

Now, let us see how the system works. In the canonical case of A-type predication, the AGR-linking condition (42) will link an NP in the external argument position to AGR. This is illustrated by an ordinary declarative sentence such as that in (44):

(44) Linguistics has warped many bright young minds.

Here, it is easy to see how (41) is satisfied: AGR is minimally governed by the P-subject Linguistics, and thus AGR-linked, by (42); both are contained in an (A-type) AGR-domain as defined by (43).

It should be emphasized that though in simple cases external θ-role assignment and predication coincide, according to this view the relation between AGR-linking and θ-role assignment is indirect. For example, examine control structures such as those in (45):

(45)a. I persuaded Bill to leave.
b. I promised Bill to leave.
c. I asked Bill to leave.

In all three cases, the P-subject is the matrix subject I, which is contained in an AGR-domain containing an infinitival complement. (Note that since untensed INFL contains no AGR specification, it will not create its own AGR-domain). However, in (45a.) it is the direct object of the matrix verb which

receives the external θ-role of the embedded predicate, whereas in (45b.) it is the matrix subject, and in (45c.) it is either the object or the subject. There is thus clearly no one-to-one correspondance between our notion of P-subject and that of "subject-of-control". Exactly the same considerations apply to the choice of controller for an adjunct predicate, as illustrated in (46):

(46)a. I ate the meat raw.
b. I ate the meat nude.
c. I talked with Bill nude.

Once again, the P-subject in all the examples in (46) is I; on the other hand, the controller for the adjunct predicate is the subject in (46a.), the object in (46b.) and either in (46c.).

It might then be asked whether our notion of P-subject bears any relation at all to more semantically-based theories of predication. In particular, does AGR-linking constrain in any way the choice of antecedent for an "unsaturated predicate" ?

The answer to this question is that there is indeed a connection, but an indirect one. To be precise, we will claim that the "missing" argument in an obligatory control structure must be contained in the same AGR-domain as the predicate which it controls. This will rule out structures such as those in (47):

(47)a. I_i said that Bill promised [to leave]$_i$

b. I told Bill$_i$ that Mary wanted [to go home]$_i$

However, within an AGR-domain, our account makes no prediction as to which NP is the controller for a given predicate, in contrast to the theories of Williams (1980), Rothstein (1983), and Culicover and Wilkins (1984), all of which attempt to incorporate the phenomenon of obligatory control into the theory of predication. Our claim is that the choice of controller within an AGR-domain is determined by a combination of the lexical properties of control predicates and various criteria of semantic appropriateness, not by structure-based grammatical principles; according to this view, the "Minimal Distance Principle" first invoked by Rosenbaum (1967) and incorporated into most accounts of control put forward since is not a syntactic principle, but an epiphenomenal tendency based on the interaction of semantic, lexical and pragmatic factors.

In Chapter 4, we will extend the theory of predication introduced here from "A-type" (θ-related) predication to "A'-type" (discourse-related) predication; it will be argued there that A'-type predication can provide a principled basis for an account of SAI-type inversion rules.

2.34 Summary

To summarize: we have made some rather major alterations to the orthodox GB perspective on θ-theory. In particular, we have relaxed the θ-criterion to permit NPs to receive more than one

θ-role, which in turn has led us to dispense with the ungoverned empty pronominal PRO; we have distinguished between internal θ-role assignment, which takes place under θ-government, and external θ-assignment, which does not; and we have introduced an entirely new concept of predication, which constrains θ-role assignment only in a very "indirect" manner.

2.4 Case-Theory

2.40 Argument Visibility

Case-assignment in GB theory is an abstract relation that may or may not show morphological reflexes. Lexicalized NPs must have Case; in Chomsky (1981) this effect is achieved through a filter applying on the PF side of the grammar:

(48) *$[_N \alpha]$, where α has a phonetic matrix, if N has no Case.

Bouchard (1984) extends the intuitive content of this filter into a more general Principle of Lexicalization, which states that there is a one-to-one relationship between Case and lexicalization.

On the other hand, several researchers, including Chomsky himself (1981, Chapter 6), as well as Brody (1984) and Levin & Massam (1985) have taken the PF role of Case to be secondary to what is generally known as Argument Visibility (henceforth A-visibility), an LF licensing condition which roughly speaking

states that in order to be "visible" for LF interpretation an A-chain (that is, a series of coindexed categories linking argument positions) must bear Case. Most versions of A-Visibility are complicated by the need to include PRO as the head of an argument chain; however, since we have dispensed with PRO, we can state the condition quite simply as in (49):

(49) An A-chain must be associated with Case.

This is rather a minimal formulation, and may have to be supplemented: see Massam (1985, Chapter 5) for a detailed discussion, and Chomsky (1986a.) for a formulation (the "Chain Condition" which forces Case to "head" a chain).

As mentioned above, we divide Case-assignment into two types, "internal" and "external", just as we have divided θ-role assignment. Let us examine these in turn.

2.41 Internal Case-Assignment

Internal Case-assignment takes place in a Canonical Government Configuration, though it is less restricted than θ-government, being able in certain circumstances to "penetrate" a complement XP in order to (exceptionally) Case-mark its specifier. Internal Case is typically assigned by the lexical Case-markers [V] and [P], and is subject to an adjacency condition, which is itself subject to parametric variation: in English, a Case-assigner must be strictly adjacent to the NP which it Case-

marks, as exemplified by the contrast between (50a.) and (50b.) below:

(50)a. I would like John to leave.

b. I would like very much *(for) John to leave.

The adverbial phrase very much inserted between the Case-assigner like and its Case-recipient John disrupts the Case-marking relation between the two, necessitating insertion of the Case-assigner for. In French, this adjacency requirement is not as strict; for example, an adverb can intervene between a verb and its direct object: compare (51a.) below with its ungrammatical English equivalent in (51b.):

(51)a. J'ai lu attentivement le livre.

b. *I read carefully the book.

For a thorough discussion of Case-adjacency, see Stowell (1981).

Languages may also differ in directionality of Case-assignment: a very productive area of research within GB theory has explored the relations between word-order typology and Case-assignment (Koopman 1984, Travis 1984). In particular, Case-theory provides a potential solution to the problem of languages which show VSO surface word-order, such as Celtic languages (Sproat 1983, McCloskey 1979), Arabic (Elesseily 1985), Tongic languages of Polynesia (Massam 1985)

and the Tsimshian languages of the N.W. pacific coast (Belvin 1984). Within a theory such as GB which uses phrase-structure configurations to differentiate internal arguments (complements) from external arguments (subjects) such languages pose an obvious problem: since there is no (surface) VP, there is no way, based on (surface) configurations, to differentiate subjects from complements. Within the theory presented here, the problem remains, even though, rather than deriving Case-marking environments from phrase-structures, we are using Case-marking to induce phrase-structures.

Case-theory provides a solution to this problem in the following way. First of all, it is assumed that VSO languages are underlyingly SVO, with a configurationally defined VP. Second, it is observed that the canonical direction for Case-assignment (hence, for Case-government) in VSO languages is to the right (see Greenberg 1963 for cross-linguistic evidence supporting this claim). Next, it is assumed that INFL assigns nominative Case in the direction of canonical Case-assignment - i.e., to the right. Finally, it is observed that INFL in VSO languages is syntactically bound to (tensed) V. In that case, the subject can only get (nominative) Case if the verb moves into a pre-subject INFL position from its underlying position in VP: hence, VSO surface word-order.

Note that the existence of non-configurational languages (if indeed they do exist: see Chomsky 1981, Saito 1985, Hale 1983) necessitates the introduction of some "assign Case" mechanism which cannot be tied to lexical or structural assignment of

Case. We will leave this as an open question, pending further research; for speculations on the learnability of such languages, see Pinker (1984).

2.42 External Case

External Case-assignment may be distinguished from internal Case-assignment in at least the following ways:

(i) In direction of Case-assignment. In English, this is obvious: the two external Cases, nominative and genitive, are assigned to the left, whereas, as we have discussed above, the canonical direction of Case-assignment is to the right. Cross-linguistically, however, this distinction may not always hold; the Case-based account of VSO languages, for example, assumes that INFL assigns nominative Case in the canonical direction of Case-assignment. Since at least some VSO languages show (configurationally instantiated) subject/non-subject asymmetries, it appears that direction of Case-assignment cannot serve in a cross-linguistic context as a sufficient condition for differentiating external from internal Case-assignment.

(ii) In adjacency parameters. As discussed above, in English there is a strict adjacency requirement between an internal Case-assigner and the element to which it assigns Case. As pointed out by Travis (1984), the adjacency requirement between

an external (nominative) Case-assigner (i.e., INFL) and the NP to which it assigns Case is not as strict: adverbial and parenthetical material may intervene between the subject NP and INFL quite freely:

(52)a. John probably did not realize how late it was.

b. John, I think, did not realize how late it was.

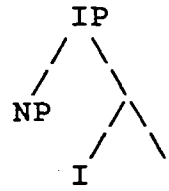
I do not know whether a similar contrast holds in VSO languages; if it did, we would predict that in tensed clauses adverbs etc. would be able to occur both between the fronted V and the (nominative) subject NP and between the subject and complement NPs, but in infinitival clauses, adverbial material would <u>not</u> occur between a (non-fronted) verb and its Case-marked complements.

(iii) In level of application. In conventional accounts, external Case is assigned at S-structure, since NP-movement, a syntactic process, is forced by external Case requirements, whereas internal Case is assigned along with internal θ-roles at D-structure. We will adopt a rather different view, claiming that both external and internal Case-asignment take place at <u>NP-structure</u>, but that the latter type involves "virtual" movement from a D-structure position (virtual, because D-structure is not a phrase-structural representation; see 2.5 below for details).

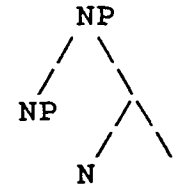
(iv) In obligatoriness. Following Levin & Massam (1985) (henceforth LM), we assume that Case-assigners must assign Case. However, whereas internal Case-assigners only assign Case if they assign a θ-role (even if, as in ECM structures, the θ-role and the Case do not match), external Case-assigners assign Case whether or not a θ-role is assigned. This means that nominative Case must be assigned whenever INFL is present (a view which will be slightly modified below), and genitive Case must be assigned whenever N is present. Since we analyze determiners as realizations of genitive Case, both the complementary distribution of determiners and subjects of NP, and the obligatory presence of one or the other of them in NP, will follow, without further stipulation, from the requirement that genitive Case be assigned.

(v) In structural conditions on Case-assignment. As discussed above, internal Case is assigned in a minimal government configuration; external Case, on the other hand, is assigned in maximal government configurations, as shown below for S and NP:

(53)a.



b.



Here Case-assignment by I or N to the subject position of NP or IP (=S) takes place within the maximal projection of the Case-assigner but not under c-command: this is the structural

relation we have identified as maximal government.

Thus we see that the partition we have established between internal and external θ-role assignment extends naturally to the theory of Case-assignment. In section 2.9 we will use this parallel asymmetry to provide a model for certain aspects of language acquisition, arguing that phrase-structure configurations may be derived from the interaction of asymmetries in Case and θ-role assignment. Before leaving this section, however, we will examine an important parametric variation in Case-assignment, based on ergative/absolutive versus nominative/accusative Case systems.

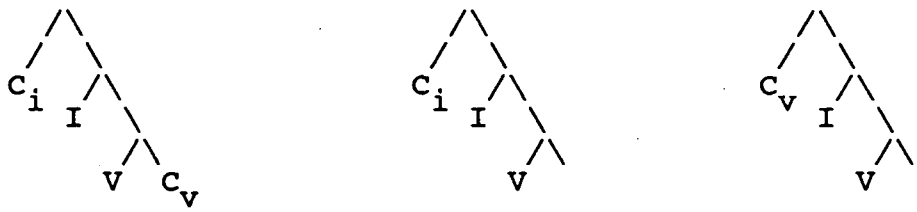
2.43 The Ergative Case-marking Parameter

There is very limited cross-linguistic variation in which Cases may be realized "externally". As discussed by LM, external Case-marking fall into one of two patterns: the nominative/accusative (N/A) pattern exemplified by English and Indo-European languages generally, and an ergative/absolutive (E/A) pattern exemplified by Tongan, Tsimshian, Inuit, and various Austronesian languages, amongst others. In a typical (surface) ergative/absolutive language, absolutive Case is assigned to the subject of an intransitive verb and the object of a transitive verb; ergative Case is reserved for the subject of a transitive verb. It is thus absolutive, an "internal" Case, which must be assigned in such languages. Adopting LM's Condition (4A), we may simply state

(54) C_x must be assigned.

where C_x is nominative in N/A languages and absolutive in E/A languages. This difference is schematized below; following LM, we will refer to NPs Case-marked absolutive and accusative as C_v and those marked nominative and ergative as C_i:

(55)a. b. c.



(55a.) shows us that in both N/A and E/A languages, C_i is assigned externally when C_v is assigned internally. However, the two types of language differ when C_v is not assigned internally: in N/A languages, C_v is not assigned and C_i is assigned externally, as in (55b.), whereas in E/A languages, C_v is assigned externally and C_i is not assigned, as in (55c.). While this follows from (55), it raises some non-trivial questions about the structural conditions for external Case-assignment. Assuming subject NPs to be immediately dominated by some projection of [I], Case-marking in the configuration in (b.) above will take place under maximal government, but in (c.) not even this weak version of government will suffice, since V, the Case-assigner, only maximally governs inside its maximal projection, VP. Various possible ways around this problem are conceivable: LM "allow C_x to percolate along the path I^o --

V^{O}", though the precise conditions under which such percolation takes place are not specified. We will (speculatively) pursue another potential solution, based on the different levels of internal and external Case-assignment.

First of all, we note that in fact there appear to be no (surface) SVO languages with (surface) E/A Case-marking (see the survey in B. Levin 1983): E/A languages appear to be either VSO (Tsimshian, Niuean) or SOV (Chukchi, Georgian, Basque, W.Greenlandic) or "non-configurational" (Dyirbal, $Yidin^{Y}$, Warlpiri). Now, as we have mentioned, most GB theorists (including LM for Niuean) derive VSO order from an underlying SVO order. Thus the absence of E/A Case-marking in SVO languages is not at D-structure, but at S-structure. This suggests that in order to assign absolutive Case to the subject of an intransitive sentence, verb-movement must take place. This makes sense in terms of the canonical direction of Case-assignment, which, as mentioned above, is rightward in VSO languages. However, E/A VSO languages differ from N/A VSO languages in that it is C_i that is uniformly assigned by a fronted, inflected verb in the latter but either C_i (in transitive clauses) or C_v (in intransitive clauses) in the former.

This in turn suggests that the fronted complex constituent V/INFL (which will be the subject-matter for much of Chapters 3 and 4) differs in Case-assigning properties in N/A and E/A languages. Continuing to accept the basic premises of the LM account of Case-marking, we will assume that Case-assigners

have a lexical feature C_x, that INFL, being non-lexical, is automatically specified C_i, and that in E/A but not in N/A languages every verb is specified C_v.

Now, let us assume that V-movement is actually movement of V/INFL (see Sproat 1983 for arguments that this is the case in Welsh). In transitive sentences of both N/A and E/A VSO languages and in intransitive sentences of N/A VSO languages, the verb will assign its Case(s) internally, merge with INFL, which will be specified C_i, move to S-initial position, and assign C_i. On the other hand, in E/A intransitive sentences, the verb will be unable to assign C_v internally, which then must be assigned externally, by (54). Thus, when the verb collapses with INFL, it will still have C_v attached. INFL will itself be automatically associated with C_i, but since C_i need not be assigned, it will be overruled by (54), and the fronted V/INFL constituent will assign absolutive Case to the subject position.

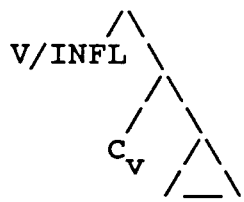
This account has several advantages over that of LM, on which, of course, it is based. It explains an otherwise mysterious gap in the distribution of E/A Case-marking, and in fact allows us to simplify (54) to the minimal formulation in (56):

(56) Ergative Case-marking Parameter: C_v must be assigned.

There is no longer any need to specify that in N/A languages C_i must be assigned, since this follows from the non-lexical

status of INFL; C_i is in fact the "elsewhere Case" and is assigned whenever it is not over-ruled by the ergative parameter or by lexicalized "quirky" Case, as in Icelandic, German, or Russian (see Pesetsky 1982, Andrews 1982). And most importantly for our purposes, Case-assignment now takes place in E/A intransitive sentences under government from the fronted verb, which is either adjoined to S or occupies a COMP position. In fact, the relevant type of government is minimal rather than maximal Case-government, as can be seen from (57):

(57)



This is in fact an "Exceptional Case-marking" type of configuration. Absolutive Case may be assigned to external arguments, but only under minimal government. Ergative Case, on the other hand, may be assigned under maximal government, exactly like nominative Case in N/A languages. We thus retain the important distinction that "lexical" Case (C_v) is assigned under minimal Case-government, whereas "non-lexical" Case (C_i) is assigned under maximal Case-government.

This concludes our overview of Case-theory. In the next section we will move onto X-bar theory, arguing that many aspects of phrase-structure can be derived from the interaction of Case- and θ-theory.

2.5 X-bar Theory and Categorial Features

2.50 X-bar Theory and Restrictiveness

Since its re-introduction in Chomsky (1970), X-bar theory has been widely accepted as a set of constraints on the phrase-structure component of the grammar. The main motivation behind its introduction was to capture the generalization that in natural languages phrases are for the most part endocentric; for example, VP below necessarily contains a V, NP an N and PP a P:

(58) $[_{NP}\text{The}[_{N}\text{explorer}]][_{VP}[_{V}\text{stood}][_{PP}[_{P}\text{on}][_{NP}\text{a}[_{N}\text{snake}]]]]$

Such phrases are "projections" of their lexical heads, with the same external distribution: the largest phrasal expansion of this type is known as a maximal projection.

There has been much debate (and confusion) over several issues in X' theory; Pullum (1985) offers a critique in which he identifies seven criterial properties of X'-grammars. These are

(59) a. Maximality
b. Optionality
c. Lexicality
d. Lexical disjointness
e. Uniformity
f. Succession
g. Centrality

Maximality refers to the condition that all non-heads in a phrasal expansion be maximal projections; Optionality to the condition that all non-head expansions be optional; Lexicality to the condition that the categorial status of a projection be the same as that of its head; Lexical Disjointness to the condition that if two projections differ in categorial status, then their heads will also differ; Uniformity to the condition that all phrases have the same "bar-level"; Succession to the condition that ensures that if a phrase X has n bars, then its expansion will include X with n-1 bars; and Centrality to the condition that the "start symbol" in the grammar be projected from some lexical category.

Not all of these conditions (known collectively as the MOLLUSC conditions) have the same status; as Pullum shows, some are empirically inadequate and some have no empirical content at all. In fact, Pullum makes use of some work on the formal properties of X-bar grammars to show that the entire set of Context Free Languages may be generated by grammars obeying the MOLLUSC conditions, as long as empty nodes are permissible. Though we do not agree with Pullum that this deprives X-bar theory of any empirical content (the study of grammars is no longer as concerned with weak generative capacity as it is with strong generative capacity, and it is properties of the latter which X-bar grammars are designed to describe) we concur with his criticisms of the exaggerated claims made on behalf of X-bar theory in the literature.

In order to meet some of these criticisms, we will "turn X-bar theory on its head", arguing in the next sections that many properties of X-bar grammars are not primitive but derived from the interaction of other more fundamental principles. By doing so, we hope to avoid the circularity implicit in accounts of X-bar syntax which define concepts such as "head" and "projection" in terms of each other; and ultimately, we hope to lay the foundations for a realistic account of the acquisition of phrase-structure.

This account is similar in many ways to that independently proposed by Speas (1986). I have tried to incorporate some of her insights into the version adopted here, in particular the idea that the succession property of X-bar grammars can be accounted for independently through the notion of "functional saturation", and her argument that percolation principles should be derivable from more primitive components of the system, just like phrase-structure rules. It seems striking to me that Speas' work, based on purely syntactic considerations, and my own, based on the perspective of epistemological priority in language acquisition, should converge on a very similar approach to phrase-structure representation; I hope that such convergence reflects the beginning of a genuine rapprochement between these two points of view.

Our account will consist of eight sections. In 2.51, we will look at X-bar theory in terms of epistemological priority; in 2.52, we will discuss basic well-formedness conditions on tree-structures; in 2.53 we will introduce a revised categorial

feature system, based on categorial association; in 2.54, we will propose a series of licensing conditions governing the mapping of various types of categories onto trees; in 2.55 we will explore one set of predictions made by this model, with respect to adjectival constructions; in 2.56 we will present a set of feature-percolation principles governing the mapping of syntactic categories onto tree-structures; in 2.57 we will suggest how these principles might be derived from the interaction of the licensing conditions; and in 2.58 we will re-examine the resulting remodelled theory in the light of Pullum's original criticisms.

2.51 X-bar Theory and Epistemological Priority

Let us take a conventional X-bar schema such as that suggested by Chomsky (1986b.):

(60)a. X' = X X"*

b. X" = X"* X'

Here, X"* stands for zero or more occurrences of some maximal projection and X = X^{0}. Order is parameterized; the ordering in (60) is that for English.

A schema such as that in (60) is based on the following proposals of Stowell (1981), which in turn follow from those originally put forward by Chomsky (1970):

(61)a. Every phrase is endocentric.

b. Specifiers appear at the X" level; subcategorized complements appear within X'.

c. The head always appears adjacent to one of the boundaries of X'.

d. The head term is one bar-level lower than the immediately dominating phrasal node.

e. Only maximal projections may appear as non-head terms within a phrase.

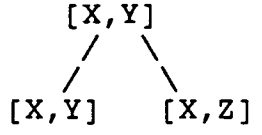
Cast in terms of Pullum's MOLLUSC conditions (see 2.50 above), principles (61a.) and (d.) instantiate Lexicality and Succession, (61e.) is a statement of Maximality. (61c.) is a separate condition, labelled "Peripherality" by Pullum, and (61b.) is not a formal condition on X-bar grammars at all. In principle, most GB versions of X-bar theory also observe Uniformity (as in the schema in (60), where the bar-level of a maximal projection is uniformly two) and Optionality (where both specifiers and complements are optional), though as pointed out by Pullum, it has been considerably less easy to put such principles into practise.

Now, consider principles such as (61) in terms of epistemological priority (see 2.0). Is there any presyntactic basis for terms such as "head", "maximal projection", "bar-level", or even "specifier" and "complement" ? The answer must surely be no; just as Chomsky (1981) rejects grammatical functions, we are forced to reject X-bar theoretic terms as plausible candidates for syntactic primitives.

In fact, most of these terms are defined in terms of each

other, leading to a type of circularity which might be benign for the linguist, but from the point of view of the child attempting to "break in" to the linguistic system, must be considered vicious. As an example of such circularity, let us examine the notion "head". Suppose we take the definition of "head" to be the set of categorial features which instantiates the Maximality, Peripherality, Lexicality and Succession conditions, as embodied in Stowell's schema by (61a., c., d. and e.) In other words, the head of a phrase XP with bar-level n will be a non-maximal, peripheral category immediately dominated by XP which agrees in all categorial features with XP save bar-level and which has n-1 bars. Thus in a phrase with the feature specification [X,Y, bar-level ^{2}] a daughter with the feature specification [X,Y, bar-level ^{1}] will automatically count as the head of the phrase, as in (62) below:

(62)



This is all straightforward, but entirely circular, since there is no independent definition of "phrase". We might just as well say that a phrase X^{n} is a projection of a peripheral, non-maximal category X^{n-1} which immediately dominates X^{n-1} and agrees with it in all categorial features save bar-level. But this merely rephrases the question of how a child calculates bar-level in the first place.

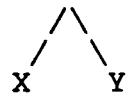
In order to avoid this circularity, we need to define X-bar

theoretic terms using other, epistemologically primitive concepts. In the model developed here, this is built into percolation principles which map Case- and θ-relations onto trees. Before turning to these principles, however, the ground must be prepared in two ways; first of all, some basic principles of tree-construction must be assumed, and second, a theory of categorial features must be developed. This will be the aim of the next two sections.

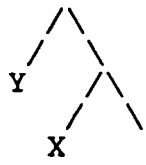
2.52 Principles of Tree-Construction

Since X-bar grammars are of course a subset of phrase-structure grammars, certain very basic conditions on the well-formedness of PS-grammars also apply to X-bar grammars. Following the approach adopted by theoreticians working in Generalized Phrase Structure Grammar (see in particular Gazdar et al. 1985) we will assume that all trees are built from "local" - i.e., one-storied - trees and that non-local trees are licensed iff all their constituent local trees are licensed. Unlike GPSG, however, we do not require that all rules apply to local trees. Thus our definition of minimal government involves the local tree configuration in (63a.) but our definition of maximal government the non-local tree in (63b.):

(63)a.



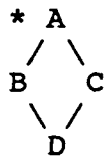
b.



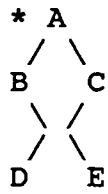
It will be observed that the distinction between conditions which apply to local trees and those which apply to non-local trees reinforces once more our distinction between "internal" and "external" syntactic processes.

Treating all non-local trees as constructed of local trees has several advantages over other approaches to phrase-structure. In particular, it eliminates the need to stipulate separately two almost universally accepted constraints on phrase-structure. These are the "unique mother" constraint illustrated in (64a.) and the "no crossing lines" (graph deformation) constraint, illustrated in (64b.):

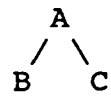
64a.



b.



c.



Neither (64a. nor b.) are constructed from acceptable local trees, which all have the form in (64c.) (abstracting away from the issue of how many branches are to be permitted per tree).

2.53 Categorial Features and Categorial Association

Most GB work, as is frequently pointed out (see for example Gazdar et al. 1985, p.19) virtually ignores the role of categorial features, generally assuming without comment the schema adopted in Chomsky (1970), which accounts only for the "major" categories A, N, V, and P:

(65) A = [+N, +V]
N = [+N, -V]
V = [-N, +V]
P = [-N, -V]

I would like to offer an alternative, similar at least in spirit to the work of Higginbotham (1986) and Speas (1986). I will distinguish three types of category: θ-heads, which includes [N], [V], and [A]; G-heads, which includes INFL and DET(erminer); and C-heads, which includes complementizers and prepositions. These three types of category enter into particular categorial associations with each other based on the notion of "functional discharge" as elaborated by Abney (1985), Fukui and Speas (1986) and the authors mentioned above. Let us briefly discuss each type in turn.

θ-heads include "primary θ-related" categories, i.e., [N], [V]; and the "inflectionally defective" θ-head, [A], which will be discussed in 2.55.

G-heads - i.e, the categories [D] and [I] - do not assign primary θ-roles but instead contain a small set of syntactically relevant features, including those involving

tense, definiteness, and number and person agreement. G-heads enter into categorial association with θ-heads either in the lexicon, the syntax, or at PF. In Chapter 3 we will explore in detail the association between [I] and [V]. Note that unlike Higginbotham (1986), we do not require [I] or [D] to discharge any particular semantic function; in fact, we regard G-heads as quintessentially syntactic in nature. G-heads also assign external Case (i.e., nominative in S and genitive in NP).

We have included prepositions and complementizers under the heading of "C-head"; it may well be that in fact, as cogently argued by Emonds (1985), both are of the same syntactic category, [P]. C-heads are "complementizers" in the sense that they act as linkers between θ-heads and their complement and adjunct dependents; they may also assign Case, like prepositions, or "secondary" θ-roles (again, like locative or directional prepositions). In addition, they may bear G-features, such as tense and AGR. It may well be that in fact C-heads have no "intrinsic" features of their own, and must "borrow" features from G-heads and θ-heads; we will not pursue this interesting possibility here. Just as G-heads enter into categorial association with θ-heads, so C-heads enter into categorial association with G-heads.

Thus we get the following schema:

(66)

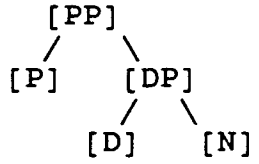
	C-heads	G-heads	θ-heads
Nominal system	P	D	N
Verbal system	C	I	V

We will assume that the heads {P, D, N} are canonically linked in the nominal system, and the heads {C, I, V} are likewise canonically linked in the verbal system. Each system will also be canonically associated with certain features: typically, these involve definiteness, specificity, and quantity in the nominal system and tense, aspect and modality in the verbal system. Non-canonical matchings are possible, but marked: thus Kwakwala allows tensed NPs, for example (Bob Levine, p.c.).

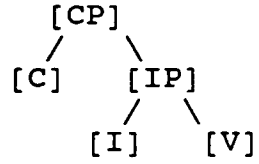
It will be observed that the category of adjective is missing from this schema; we will discuss the "defective" nature of adjectives below in 2.55.

Given the system in (66), we will get the following phrase-structure representations:

(67)a.



b.



(Note that the representations above are simplified, in that they abstract away from argument structure: see 2.56 below)

As mentioned above, we allow the possibility for morphological as well as syntactic association; in Chapter 3, we will explore the nature of various parameters relating to levels of [I,V] association, and the types of feature percolation permitted when such association takes place in the syntax. We shall have very little, on the other hand, to say about the nature of [D,N] association; see Abney (1985), Stowell (1987), for some

recent work along the lines of the approach suggested here.

2.54 Syntactic Licensing Conditions

What forces the categorial associations in (76) ? We could, of course, simply stipulate them as part of the phrase-structure component; however, this would run counter to our express aim of creating a purely derivative notion of phrase-structure. Instead, we will adopt a "licensing conditions" approach (see Chomsky 1986a.), in which conditions on interpretation at LF and PF act as filters on syntactic representations. We have in fact already introduced two sets of licensing condition: the "Case-visibility condition" in (49), repeated as (68), and the "Predicate-visibility conditions" in (41), repeated below as (69):

(68) An A-chain must be associated with Case.

(69) I. Every predicate must be contained within an AGR-domain

II. Every AGR-domain must contain a Predication Subject.

III. Every predication subject must be AGR-linked.

We will now propose the following general condition on G-heads, which we will refer to as G-visibility:

(70) G-heads must be θ-associated.

In Chapter 3, we will investigate (70) in more detail, when we examine the conditions under which [I] associates with [V]. G-heads contain G-features, including AGR-features, which in turn license both predicates and external arguments; therefore, together with the predicate-licensing conditions of (69), (70) ensures (indirect) linking between VP and its external argument. Next, we will propose the following C-visibility condition:

(71) C-heads must be G-associated.

(71) is a companion condition to (70); the two together force a particular "association heirarchy" on phrase-structure, as we will see in 2.56 below.

Note that since C-heads are not necessarily either Case- or θ- associated, they will be obligatorily present just in case C-features (such as [±WH]) are selected. Since licensing conditions are in effect a particular formulation of syntactic "recoverability", we predict that complementizers and prepositions may be deleted and inserted far more freely than categories constrained by visibility conditions. This is indeed the case: there are many cases of P-insertion (e.g, of-insertion) and complementizers can be freely deleted up to recoverability (see Lasnik and Saito 1984).

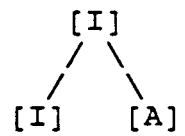
The licensing conditions in (68-71) express the interdependence of syntactic categories in a particular way. For example, one of the claims embodied in this formulation is that

subjects and predicates are only indirectly linked. Thus, assuming that a predicate is the central component of a syntactic representation, P-licensing condition (I) ensures that every predicate will be contained in an AGR-domain, which in turn ensures the presence of AGR, and by extension, INFL. By P-licensing condition (II), each AGR-domain will have a P-subject; therefore, indirectly, every predicate will be linked to a subject.

2.55 Adjectives and Feature Percolation

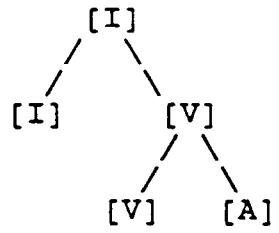
In this section, we will examine briefly the θ-head [A], corresponding to the traditional category of Adjective. [A] assigns no Case at all, and cannot be associated with INFL. In our terms, [A] cannot count as a "primary θ-head" and therefore does not meet the G-visibility condition in (70). Thus structures such as (72) will be ruled out:

(72)



Since they are "inflectionally defective", in order to assign internal θ-roles, adjectives must subcategorize for PP complements, and in order to assign an external θ-role, the dummy verb be must be inserted to bear Tense, which in turn enables AGR to assign Case to the external argument and create an AGR domain for predication, as in (73):

(73)



A special property of the "empty" θ-assigner be allows it to transmit a θ-role from an adjectival head to its external argument. Note that we are assuming that be does assign a θ-role; however, be is unique in being able to "adopt" a θ-role from its complement. The difference between be and an ordinary θ-assigning verb is best illustrated in terms of a version the theory of "θ-grids" first proposed by Stowell (1981). In Stowell's model, each verb is specified in the lexicon for the set of θ-roles which it assigns; thus an intransitive verb such as sleep will be represented as in (74a.), a transitive verb such as love as in (74b.), and a ditransitive such as show as in (74c.):

(74)a. sleep: θ_1
|
experiencer

b. love: θ_2 θ_2
| |
agent patient

c. show: θ_1 θ_2 θ_3
| | |
agent theme recipient

Adjectives, like verbs, will be associated with a θ-grid. Thus an monadic predicative adjective such as red will be

represented as in (75a.) and a dyadic predicative adjective such as <u>fond</u> as in (75b.):

(75)a. red: θ_1
|1
theme

b. fond: θ_1 θ_2
|1 |2
experiencer theme

Now, the special status of <u>be</u> lies in its ability to bear an unspecified θ-role, as in (76):

(76) be: θ_1
|1

This means that <u>be</u> can "adopt" a θ-role from a complement adjective phrase by a process of "linking" analogous to that much-used in autosegmental phonology:

(77)

In this way, we can account for both the obviously verbal status of <u>be</u> and its ability to transmit θ-roles from a complement XP to an external argument.

This completes our categorial survey. We will now turn our attention to principles of categorial feature-percolation, which determine the status of non-lexical (i.e., non-pre-

terminal) nodes.

2.56 Principles of Categorial Feature Percolation and the GREPP

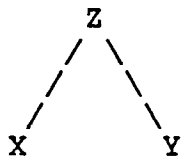
In this section, we will propose explicit feature-percolation principles, which represent a particular instantiation of the feature-association model presented in 2.53, and anticipate the discussion of the learning of phrase-structure in 2.9 below. These principles are not meant as primitives of the theory; as pointed out by Speas (1986), if we are to derive phrase-structure from the interaction of other, more primitive principles, then percolation rules should be as unnecessary as explicit phrase-structure rules. Rather, they are meant as "operating principles" which relate licensing conditions to strings to derive labelled trees.

The first question to be asked in formulating a theory of feature percolation is whether it should be "top-down" or "bottom-up". It is by no means clear that the two are not notational variants. We will assume a "bottom-up" percolation mechanism in order to mimic the language-learning process, where we will assume that the properties of lexical items are generally acquired prior to the acquisition of the properties of phrasal categories. In effect, the latter will follow from the former, given an adequate theory of percolation.

Let us then proceed to the principles themselves. Assuming for the moment that we are dealing only with local trees, we need to know in a structure such as (78) below, for any given

value of X and Y, which categorial features will percolate from the daughter nodes to the mother node Z:

(78)



(Note that (78) is unspecified for the linear order of X and Y, which is language-specific, and will depend on the CGC of the language in question.)

First of all, suppose X θ-governs Y. In that case, X will θ-mark Y, and the features of Z will be those of X. This allows us to formulate our first percolation principle:

(79) Percolation Principle I:

> Where X θ-governs Y, the categorial features of Z will be those of X.

Principle I will operate quite straightforwardly where both Case and a θ-role are assigned by the same category to the same category - i.e., by V to the internal arguments of VP. Here the categorial features of V will percolate, as predicted. In this sense, the relation between V and its θ-governed complements is the "core-case" of government, which is exactly what the notion of Canonical Government Configuration is designed to capture.

Next, let us examine situations where Case-marking does not

coincide with θ-marking. These are of two types: firstly, there is the relation between categories linked by Case-assignment (under Case-government) but not θ-role assignment; and secondly, the relation between categories θ-linked under Predication but not Case-linked. Let us examine them one by one.

First, let us consider Case-assignment without (primary) θ-role assignment. There are three main types: the relation between [I] and a nominative subject NP; that between [N] and a genitive subject NP; and that between [P] and a prepositional object. In all three cases, where X is a Case-assigner and Y the category to which Case is assigned, categorial features of Z will be those of X. Thus we might formulate Principle II as below:

(80) Percolation Principle II:

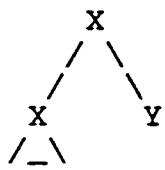
> Where X assigns Case to Y, the categorial features of Z will be those of X.

It might be thought that Principle II in fact makes Principle I redundant, since θ-government generally involves Case-assignment. However, there is a class of verbs (the so-called unaccusative verbs) which assign an internal θ-role under θ-government but do not assign Case, forcing the θ-marked NP to move to subject position. In such cases, Principle I will apply to the NP-trace under VP, but Principle II will not, since no

Case-relation exists between V and its complement. Similar considerations apply to the trace of passivized NPs.

In developing phrase-structure percolation principles we have so far concentrated exclusively on what we might call the "subcategorization set" of a phrase, which includes all elements linked by Case- or θ-relations, as well as lexically selected G-heads - in other words, the "essential ingredients" of a phrase. Let us now turn our attention to the "adjunction set" of a phrase, which will include all non-subcategorized adjuncts, modifiers and the like. Such elements typically have no effect at all on categorial structure. In other words the category dominating the adjoined/modified category has exactly the same categorial status as the category to which it is adjoined:

(81)



Here, Y is either a base-generated or syntactically moved adjunct, and X the category to which it is adjoined. The same structure will characterize adjectival modifiers and relative clauses in NP and presumably VP and S adverbials. This suggests the following principle:

(82) Percolation Principle III:

Where X is a member of the adjunct set and Y a member of the subcategorization set of a phrase Z, the categorial features of Z will be those of Y.

While the principle itself is quite straightforward, and the intuitive content of the distinction between subcategorization and adjunct sets is clear, we have not yet proposed a way to grasp this distinction using the formal apparatus available to us within the theory outlined here. In order to do so, we will at this point appeal to a much-modified and generalized version of the Revised Extended Projection Principle as discussed in section 2.11, which we will, accordingly, label the Generalized Revised Extended Projection Principle (GREPP):

(83) Subcategorization requirements must be satisfied by all phrase-structure configurations, where "subcategorization requirements" refers both to subcategorized and subcategorizing elements.

There are two main ways in which (83) differs from standard versions of the Projection Principle. Firstly, rather than referring to "all levels of the grammar" (i.e., D-structure, S-structure and LF) we have simply referred to phrase-structure configurations. This alternative formulation is probably equivalent to its predecessor within orthodox approaches to GB theory, where each grammatical level is defined in terms of a particular set of phrase-structure configurations, but not

within the approach taken here, since there is no presyntactic phrase-structure component onto which other grammatical principles are mapped, but only a set of well-formedness conditions on trees and categorial percolation principles which are themselves based on the interaction of other components. Thus "D-structure" within the model proposed here is nothing more than a set of (internal) Case- and θ- relations, with at most a partial representation in configurational terms.

The second major difference between (83) and the standard Projection Principle is that we have extended the notion of "subcategorization requirements". Thus we have included the class of subcategorizing as well as subcategorized elements in the subcategorization set of a phrase, in effect generalizing the Projection Principle from a constraint on the distribution of NPs to a general constraint on the recoverability of lexical information.

Thus formulated, the GREPP makes the required distinction between the subcategorization and adjunction sets of a phrasal category.

Next, let us turn to the percolation relations holding between G-heads and θ-heads. We will formulate the following principle:

(84) Percolation Principle IV (first version):

Where X is a G-head and Y a θ-head, and X and Y are in a CGC, the features of X will percolate to Z.

Principle III will apply to the relation between [V] and [I] and also to the relation between [D] and [N].

Next, let us turn to the relation between G-heads and C-heads. Here, we formulate the principle as in (85):

(85) Percolation Principle V (first version):

> Where X is a C-head and Y a G-head, and X and Y are in a CGC, the features of the C-head X will percolate to Z.

It should be noted that as stated, principles IV and V are virtually identical. This suggests that we might collapse them into a single percolation principle incorporating a <u>feature percolation heirarchy</u>, as formulated in (86) below, which replaces both IV and V:

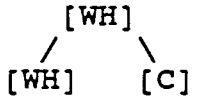
(86) Percolation Principle IV (revised version):

Where X and Y are in a CGC, no Case or θ-relation holds between them, and both are part of the subcategorization set of Z, the following heirarchy determines which features will percolate:

(a) C-features of X and Y will percolate to Z
(b) G-features of X and Y will percolate to Z
(c) θ-features of X and Y will percolate to Z

However, one problem remains: we have not yet dealt with the relation between a complementizer and its "A'-subject" - i.e., a WH-phrase. None of the principles we have introduced so far will account for the percolation relations holding between the two, since [WH] is not (directly) involved in Case or θ-relations and it does not appear to be either a G-head or a C-head in a CGC with its "complement". Yet [WH] is available as a lexically selected feature: only certain verbs subcategorize for WH-complements, and when they do, no value for [±TENSE] may be selected (see Stowell 1981, Speas 1984). Now, as we have already mentioned in 2.3 above, the distribution of [WH] is limited to "SPEC C" environments; this suggests that we grant it quasi-categorial status as a C-head, in which case Principle IV will apply to yield a configuration such as the following:

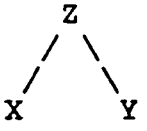
(87)



Assuming that only heads are available for feature-selection, this correctly predicts the facts mentioned above.

Let us summarize. In structures such as (88)

(88)



the following principles of categorial feature percolation hold:

Percolation Principle I:

Where X θ-governs Y, the categorial features of Z will be those of X.

Percolation Principle II:

Where X assigns Case to Y, the categorial features of Z will be those of X.

Percolation Principle III:

Where X is a member of the adjunct set and Y a member of the subcategorization set of a phrase Z, the categorial features of Z will be those of Y.

Percolation Principle IV

Where X and Y are in a CGC, no Case or θ-relation holds between them, and both are part of the subcategorization set of Z, the following heirarchy determines which features will percolate:

(a) C-features of X and Y will percolate to Z
(b) G-features of X and Y will percolate to Z
(c) θ-features of X and Y will percolate to Z

Perhaps at this point an example of the operation of the percolation principles might be helpful. Let us take the sentence in (89):

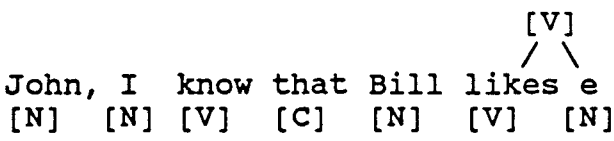
(89) John, I know that Bill likes.

Now, assuming that all lexical categories have been correctly identified, we will get the "lexical string" in (90):

(90) John, I know that Bill likes.
[N] [N] [V] [C] [N] [V]

Again, assuming that the subcategorization properties of the various lexical items are known (since this is important to the operation of the GREPP, which in turn determines the distribution of empty categories,) we can begin to apply the principles themselves, beginning with the most embedded phrase, since we are assuming a bottom-up procedure. First of all, by Principles I and II, and by the GREPP, we construct the local tree in (91):

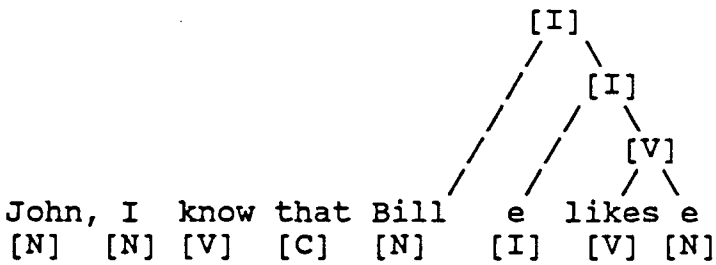
(91)



Next, we come to the Case-marked NP Bill. By the GREPP, there must be an element which assigns Case to this NP; since the NP

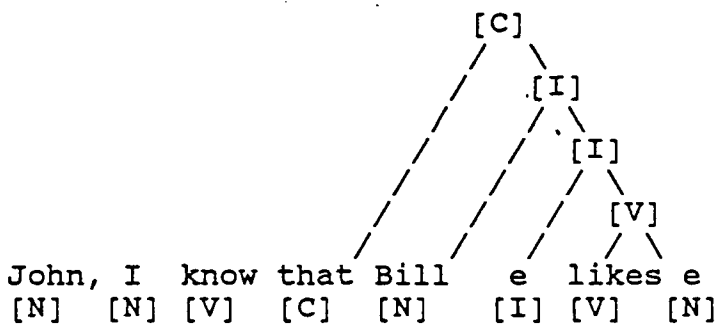
receives its Case in a non-canonical (external) government configuration, that element must be a G-head; since the G-head associated with [V] is [I], the missing element must be [I], and by Principles IV and II the relevant structural representation will be as in (92):

(92)



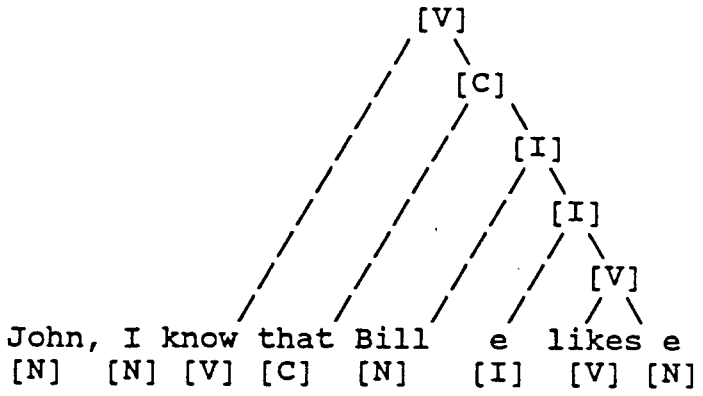
Next, by Principle IV, we will get the structural representation in (93):

(93)



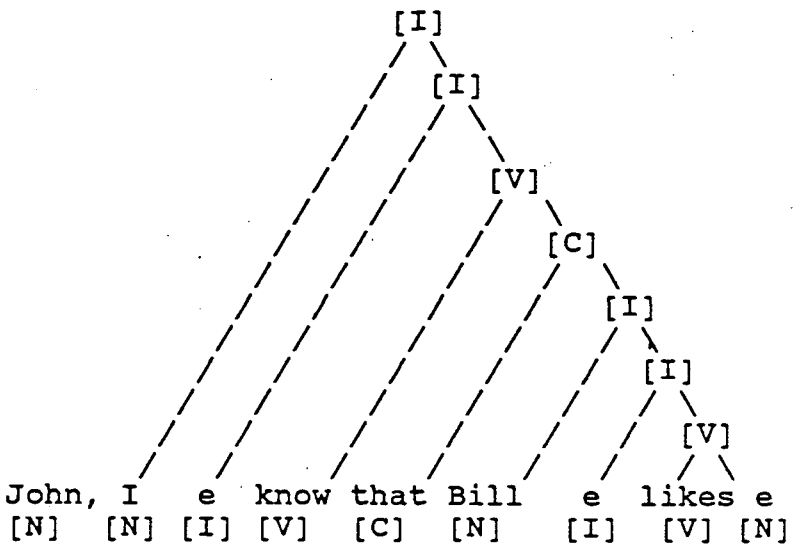
By Principle I, we then construct (94):

(94)



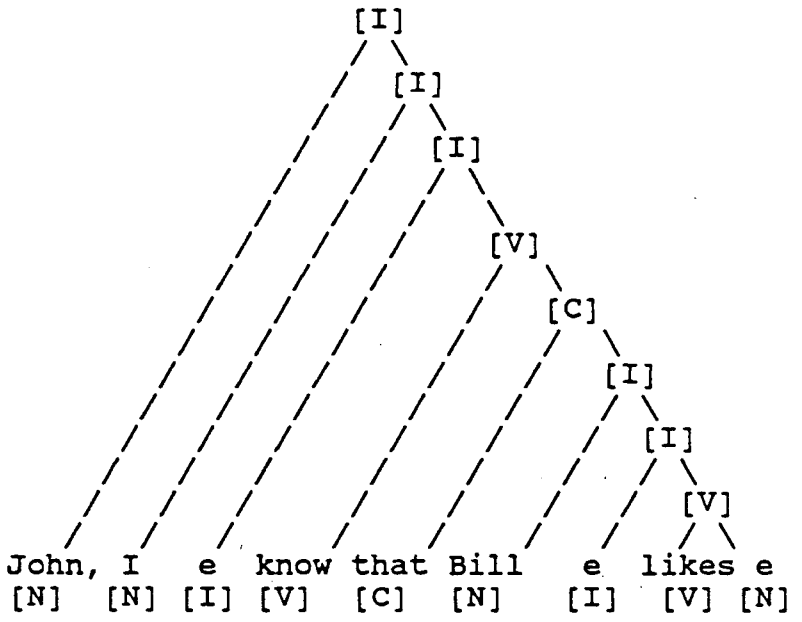
Next, by the GREPP, Principle IV, and Principle II, we get (95):

(95)



Finally, by Principle III, we end up with the phrase-structure representation in (96):

(96)



2.57 Deriving Percolation Principles

So far, we have proposed an explicit set of Percolation Principles, which, as can be seen from the examples above, can be used to generate labelled trees from strings. However, at the outset of this section, we suggested that such principles were in fact no more than "realization rules" for the mapping of more fundamental licensing conditions onto trees. In this section, we will adopt a version of the theory of "saturation" as developed by Higginbotham (1985, 1986) and Speas (1986) in order to derive percolation principles from licensing conditions. We will differ from Higginbotham's approach (and also from the very similar approach of many linguists working in the Montague grammar framework) by assigning no special semantic status to saturation; rather we will treat θ-roles as

syntactic features on a par with Case, AGR(eement) features, and so forth.

The basic idea of saturation (due to Bar Hillel 1953) is that a process of quasi-arithmetical cancellation characterizes the "discharge" of functional roles. Thus, a verb which takes two thematic arguments is "cancelled out" by categorial combination with two NPs bearing appropriate θ-roles. Higginbotham (1986) generalizes θ-saturation to all major syntactic categories (claiming that INFL, for example, is associated with an "event" θ-role), and Speas (1986) extends the idea further, claiming that saturation of a generalized notion of "Kase" plays an equivalent role to θ-saturation for non-lexical "functional" categories.

We will adopt a version of this framework in which each set of syntactic features $\{F_i...F_n\}$ associated with a particular category must be assigned (or "discharged") as a condition on well-formedness. Discharge of a particular feature F_j leads to syntactic "closure" for F_j. Thus a predicate which has assigned all of its (internal) θ-roles will be θ-closed, a Case-assigner which has assigned all of its Cases will be Case-closed, and so on.

It is easy to see how this approach will account for Percolation Principles I and II, which each involve a relation between an "open" and a "closed" category; in both cases, the features of the open category will percolate to the dominating category. Percolation Principle III, which involves the relation between members of the adjunct and subcategorization

sets of a phrase, and Percolation Principle IV, which involves categorial association, are less easy to deal with. In fact, we have little to say about how principle III may be reduced to independent principles; note that as stated, it is in any case a consequence of the distinction, made by the GREPP, between the adjunct and subcategorization sets of a phrase. See Speas (1986) for further comments.

As for Principle IV, note that the G-visibility principle in (70) and the C-visibility principle in (71) will independently force the categorial associations ({I,V}, {C,I}, {D,N}, {P,D}) and thus account for the "percolation heirarchy" in Principle IV. In terms of saturation, we can think of the functional G-head and C-head categories as requiring θ-head categorial features for closure; in this sense, they "assign categories", as opposed to assigning Case or θ-roles (though of course, they may assign both).

This has been a cursory and tentative discussion, and much remains to be worked out; nevertheless, it seems to me that at least in principle we have shown that the percolation principles introduced in the last section may be derived from other, more fundamental aspects of the system.

2.58 Categorial Feature Percolation and the MOLLUSC Conditions

We have seen that in tandem with the GREPP, categorial percolation principles can meet the Lexicality requirement on phrase-structure rules (including of course, X-bar schemata).

However, Lexicality is just one of the seven MOLLUSC conditions identified by Pullum as characteristic of X-bar grammars, and is in and of itself, as he points out, "of no consequence". It is only together with other conditions that Lexicality has an effect, and it is to these other conditions that we now turn.

Let us review once more the seven MOLLUSC conditions of Pullum (1985), repeated here as (97):

(97)
a. Maximality
b. Optionality
c. Lexicality
d. Lexical disjointness
e. Uniformity
f. Succession
g. Centrality

So far we have only addressed the issue of Lexicality. Of the remaining six conditions, Lexical Disjointness is really only of technical (mathematical) interest, and in a theory such as the one presented here, where the features of every non-lexical node are projected from those of a lexical node, Centrality is automatically satisfied. This leaves us with conditions (a., b., e., and f.). Of these, by far the most important is Succession, which as formally stated by Pullum is as in (98) below:

(98) If X^k --> W is in R, then W = YX^{k-1} for some Y, Z in $(V_n \text{ union } V_t)^*$.

Here R refers to the set of rules of a given grammar, V_n refers

to the set of non-terminal nodes, V_t to the set of terminal nodes, and (V_n union V_t)* to the set of all strings, of length zero or greater, formed from the set formed by the union of the sets of terminal and non-terminal vocabularies. In other words, the bar-level of a head is one less than the bar level of its mother.

The following question then arises: in what sense can a system such as ours based on principles of feature percolation be said to obey the Succession Condition ?

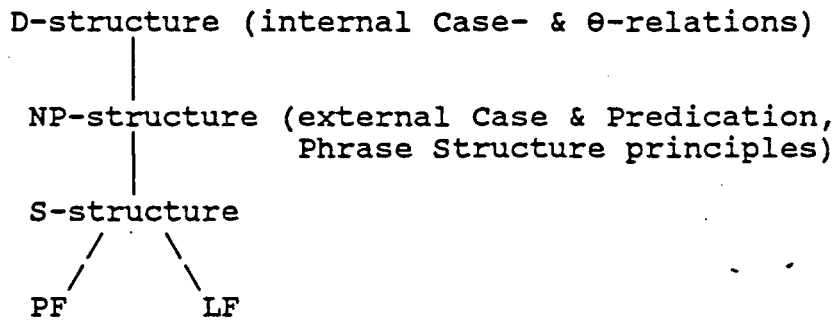
It is not at all obvious that it should. Pullum (1985) in fact argues against Succession, claiming that in the normal case different projections of a category have the same bar-level. Note that in the system adopted here we can refer to levels of "closure", rather than bar-level; thus a Case-closed V will be equivalent to a "small" VP, and so forth. The point is, we do not need to make explicit reference to bar-level in order to pick out "intermediate" projections, even if we in fact need to do so.

It will be observed that Succession (or its "degree of saturation" equivalent) must apply after categorial feature-percolation principles but before rules which make crucial reference to "intermediate" projections. This gives us two options: we might abandon the concept of syntactic movement altogether, which would bring our model closer to those based entirely on principles of feature-percolation, notably Generalized Phrase Structure Grammar (see Gazdar et al. 1985); or alternatively, we might assume an intermediate syntactic

level between D-structure and S-structure. We will take the latter option, adopting from van Riemsdijk and Williams (1981) the level of NP-structure, the output of NP-movement and the input to WH-movement. Interestingly enough, the arguments in that paper for such a level are based on quite different considerations than those discussed here, most notably problems with LF-reconstruction rules; that examination of such different grammatical phenomena should converge on the need for a level of NP-structure seems strong evidence for its adoption.

We are thus contemplating a rather different grammatical architecture from that introduced in 2.1 above; this revised model is presented below:

(99)



Within this model, the rule of "NP-movement" no longer applies to phrase-structure configurations, since it takes place prior to the construction of phrase-structure. This means we are adopting a "lexical" approach to NP-movement, though the GREPP ensures that at NP-structure empty categories appear as the "traces" of NP-movement, which are subject to the binding conditions (see 2.6 below) "as if" there had been movement. In fact, it may be possible to dispense with D-structure

altogether, which has a much-reduced status even within orthodox GB theory and in this model seems entirely reducible to information already available in the lexicon. See Culicover & Wilkins (1984, Chapter 1) for relevant discussion of this issue.

Let us then turn to the remaining three MOLLUSC conditions, Maximality, Optionality, and Uniformity. Within the theory adopted here, none of these three conditions has any theoretical status; as Pullum (1985) shows, this is just as well, since they are neither theoretically interesting nor empirically supported.

Maximality, the condition that all non-head constituents be maximal projections, is trivially satisfied in the theory presented here, since a maximal projection is no more than the highest percolation of a set of categorial features.

Optionality may be dispensed with altogether, since the GREPP ensures that the subcategorization set of a phrase is obligatorily generated, while the adjunct set is optional.

Finally, Uniformity is rejected outright in the theory adopted here; there is no theoretical reason to retain it, as Pullum points out, and the best arguments against its empirical success are those which attempt to argue for it, notably the extremely ad-hoc analyses in Jackendoff (1977) which include, for example, a three-level "Modal-phrase" where only one level is ever lexically realized.

This concludes our reformulation of the phrase-structure component of GB theory, which, as we have seen, entails several

important consequences for the architecture of the model. Obviously, the discussion presented here has been somewhat cursory; in subsequent sections of this chapter (notably 2.9) and in subsequent chapters, we will further explore some of the cosequences of "turning phrase-structure on its head." Let us now turn, however, to the remaining subcomponents of the theory. Here our discussion will be somewhat briefer.

2.6 Binding Theory

The GB approach to binding makes a systematic distinction between the three types of nominal elements found in A(rgument) positions: anaphors, pronominals, and R(referring)-expressions. A-positions are those which (potentially) receive a primary θ-role, i.e. complement positions plus subject. As stated in Chomsky (1981, p.188), there are correspondingly three parts to the binding theory:

(100) (A) An anaphor must be bound in its Governing Category.
(B) A pronominal must be free in its Governing Category.
(C) An R-expression must be free.

A few definitions are in order. A nominal is bound if it agrees in features with a (minimally) c-commanding coindexed antecedent, A-bound if that antecedent is in an A-position, A'-bound otherwise. For the purposes of the LGB version of the binding theory, bound means locally A-bound (though there have

been numerous attempts to generalize binding to include A'-binding: see for example Aoun 1985, Bouchard 1984). Governing Category is defined as below (Chomsky,op. cit. p.211):

(101) β is a governing category for α if and only if β is the minimal category containing α, a governor of α, and a SUBJECT accessible to α.

Further definitions: a SUBJECT is "the most prominent nominal element" (Chomsky op.cit.,p.209), that is, AGR in tensed clauses (which is assumed to have pronominal status) and subject elsewhere. This formulation is in effect an attempt to combine the effects of the Specified Subject/Opacity Condition (SSC), which blocks anaphor-binding and allows pronominal-binding in configurations such as (102a.) below, with those of the Propositional Island/Nominative Island Condition (NIC), which has the same effect in configurations such as (102b.):

(102)a. They$_i$ told Bill [to see *each other$_i$ / them$_i$]

b. They$_i$ said [that *each other$_i$ / they$_i$ would talk to Bill]

Accessibility is defined through a supplementary condition, the i-within-i condition (103):

(103) *[$_{\alpha i}$...δ_i...]

(104) α is <u>accessible</u> to β iff β is in the c-command domain of α and coindexing of (α,β) would not violate (103).

There is increasing evidence that the notion of governing category varies extensively from language to language; Manzini & Wexler (1984) have a five-valued parameter, for example:

(105) β is a <u>governing category for α</u> iff β is the minimum category containing α, a governor of α, and

a. has a SUBJECT, or
b. has an INFL, or
c. has a TENSE, or
d. has an indicative TENSE, or
e. has a root TENSE.

There is further variation in what may count as an antecedent for the purposes of the binding theory. For example, any c-commanding NP within an appropriate governing category may count as an antecedent for an English reflexive, but not for the Japanese reflexive <u>zibun</u>, which must be bound by a subject NP. This leads Manzini & Wexler to a further parametrization:

(106) A proper antecedent for α is

a. a subject
b. any element β whatsoever

Note that the parametrization is crucially ordered by the Subset Principle, discussed in Chapter 1.

One of the most interesting aspects of the GB binding theory

is that it treats overt and empty categories in the same way. This leads to the prediction that there will be empty categories corresponding to each type of overt NP, and indeed, this is precisely what is claimed. Corresponding to overt anaphors such as reflexives or <u>each other</u> are NP traces; corresponding to overt pronominals are the empty pronouns <u>PRO</u> (which we have abandoned; this has profound consequences for the binding theory, which will not be explored here) and <u>pro</u> (which is licensed by "rich" AGR); and corresponding to overt R-expressions are WH-traces, which have the status of operator-bound variables at LF.

Though conceptually elegant, this version of the binding theory leaves some loose ends: in particular, though WH-trace appears to escape the effects of the SSC, as exemplified by (107a.), it appears to be subject to the NIC, as exemplified in (107b.):

(107) a. Who_i did Bill tell the men [to talk to e_i]

b. *Who_i did Bill say [that e_i slept through an earthquake]

Chomsky's approach to this problem (RESNIC, or the residue of the NIC) is to propose an additional Principle, the Empty Category Principle (ECP), which applies to non-pronominal empty elements, and which constitutes the "Government" subcomponent of GB theory, to be discussed in 2.8 below. An alternative, proposed by Aoun (1985), is to assume that A'-binding is subject to both Principles A and C of the Binding Theory, in

which case a variable will be both an anaphor (since it must be bound within its Governing Category, though crucially A'- rather than A-bound) and an R-expression (since it must be A-free).

The locus of binding theory has been variously set at S-structure, LF, or both; current opinion (see Chomsky 1986a., for example) seems to favour LF.

2.7 Bounding Theory

Bounding theory has had a particularly important historical role in the development of the principle-and-parameter approach to syntax. The introduction of the subjacency principle (Chomsky 1973) showed that several apparently separate island constraints (for example, the Complex NP, WH-island, Left Branch, and Sentential Subject Constraints of Ross 1967) were amenable to a unified treatment, given certain assumptions about syntactic movement. Informally stated, subjacency allows movement over one but no more than one "bounding node", where NP and S are bounding nodes in English. Apparent "unbounded" movement can be reduced to the possibility of successive-cyclic WH-movement through COMP. We thus account for the ungrammaticality of (108a.-d.) below:

(108)a. *Who$_i$ did [$_S$ Bill see [$_{NP}$ the man$_j$ [who$_j$ [$_S$ e_j saw e_i]]]] ?

b. *Whose$_i$ did [$_S$ you steal [$_{NP}$ e_i book]] ?

c. *Who$_i$ did [$_{NP}$[that [$_S$ Bill saw e_i]]] upset John] ?

d. *Who$_i$ did [$_S$ you see [why$_j$ [$_S$ John upset e_i e_j]]] ?

(108a.) is an example of a Complex NP Constraint violation; (108b.) of a Left Branch Condition violation; (108c.) of a Sentential Subject Constraint violation; and (108d.) of a WH-island violation. In each case more than one bounding node has been crossed. The subjacency condition has played an important role in the renewed interest in comparative syntax and in the development of the parameter-based approach to cross-linguistic variation. In a seminal paper, Rizzi (1982) traced the absence of WH-island effects in Italian to a difference in the value of bounding nodes; in Italian S' and not S counts as a bounding node, allowing a WH-word to move directly from an embedded S to a matrix COMP (over an intervening WH-word in an intermediate COMP) without violating the Subjacency Condition. Compare (109) below with (108d.) above:

(109) Il solo incarico$_i$ [$_{S'}$che$_i$ [$_S$non sapevi [$_{S'}$a chi$_j$ [$_S$
the only charge that not I-know to whom

avrebbero affidato e_i e_j]]]] è poi finito proprio a te.
they-would-have entrusted is then ended exactly to you

"The only charge that I didn't know to whom they would entrust has been entrusted exactly to you."

Note that this type of parametric variation is quite incompatible with the Wexler/Manzini Lexical Parametrization Hypothesis discussed in Chapter 1, since the concept of bounding node makes no reference whatsoever to lexical

properties (though see Adams 1984 for a reformulation of the parameter which is "lexically reducible". This suggests that even if the LPH applies to some types of parametrization, it cannot apply to all.

Bounding theory has led to rather promising results in still another area, that of learnability theory. In proving the learnability of one type of transformational grammar, Wexler & Culicover (1980) were led to adopt various principles to constrain the hypothesis space of their abstract language learner. Amongst these principles was the Binary Principle (Wexler and Culicover 1980, p. 312-313) which states that

(110) A transformation may apply no deeper in a phrase-marker than the immediate constituents of the next B-cyclic node below the one at which the the transformation is applying.

(Here "B-cyclic node" refers to S and NP.)

Now, as Wexler and Culicover point out, the Binary Principle resembles the subjacency condition, which was proposed for quite different (empirical) reasons as a constraint on the application of syntactic rules. Wexler and Culicover argue that this type of convergence from two quite different types of linguistic enterprise lends considerable support to a subjacency/binary principle type of constraint.

In view of its apparent explanatory success, it is ironic

(though perhaps ultimately not surprising) that the subjacency condition should have been so dogged by empirical problems as to have barely achieved descriptive adequacy even in English. The following well-known example from Ross (1967) involves multiple violations of the Complex NP constraint:

(111) This is the report$_i$ [$_S$, which$_i$ [$_S$ the government prescribes [$_{NP}$ the height of [$_{NP}$ the lettering on e_i]]]]

When we look further afield cross-linguistically, the problems multiply. There is a general concensus amongst linguists working on Scandinavian languages that even a parameterized version of subjacency is too strong to allow the multiple extraction possibilities of, for example, Swedish, which allows violations of <u>all</u> the Island conditions mentioned above (see in particular Engdahl 1979, Maling and Zaenen 1982). This has led to the significance of subjacency (and bounding theory in general) being down-played by GB theorists in recent years; a typical comment is made by Chomsky (1986b., p.37) who refers to Rizzi's work on the parametrization of subjacency as "low-level parametric variation that receives little support in evidence available to the language learner."

A further source of contention in bounding theory has been the level at which subjacency applies. Originally formulated as a general constraint on syntactic processes affecting any two terms α and β, it has been narrowed down in GB theory to a constraint on the syntactic application of the rule "move α." The important work of Huang (1982) has been generally taken

to show that subjacency does not apply to LF-movement. For example, in certain types of multiple WH-constructions (first discussed by Baker 1970), there is a wide-scope reading for an embedded (non-echo) in-situ WH-phrase which, assuming LF WH-movement, could only be derived by a WH-island violation:

(112) Who saw where John found what ?

Here the relevant, subjacency-violating reading is the "distributed" one, where a possible answer might be: "Bill saw where John found the gold watch, Mary saw where John found the diamond cuff-links..., etc." An LF representation for this type of reading would look something like (113) below:

(113) For which X, X a person, & for which Y, Y a thing, X saw where John found Y.

This in turn can be derived from the (syntactic) LF representation in (114):

(114) [what$_j$ [who$_i$ [e_i saw [where$_k$ [John saw e_j e_k]]]]

Here the WH-island condition, and thus subjacency, has been clearly violated.

Now, if subjacency does not apply to LF representations, it may either apply as a constraint on syntactic movement (the mapping from D-structure to S-structure in most versions of GB

theory, and from NP-structure to S-structure in ours) or on the resulting S-structure representation. Evidence on this issue is much harder to come by than vitriolic rhetoric; Chomsky has recently (1986b.) abandoned a tortuous argument in favour of a movement analysis of subjacency (first presented in Chomsky 1982) based on the licensing of parasitic gaps, and I now know of only one argument which appears to show subjacency to be a constraint on movement rather than on an S-structure representation. This is based on Spanish data presented in Torrego (1984); the essence of the argument is that there is a V-preposing rule in Spanish which applies whenever an argument WH-phrase is moved to COMP in the syntax. Now, Spanish, like Italian, permits WH-island violations, and Torrego shows that if we adopt Rizzi's analysis of such violations (based on parametric variation in bounding nodes) then we predict that V-preposing should not take place where a WH-word has "skipped" a COMP node. Torrego maintains that this is indeed the case. Of course, it could always be argued that the same effect could be derived from an S-structure representation, as long as intermediate WH-traces triggered V-preposing. However, there is very good evidence that intermediate WH-traces in COMP are specified [-WH] (for example, for the purposes of lexical selection, they cannot count as [+WH]); thus the only plausible explanation is that subjacency is a constraint on WH-movement and not on S-structure representations. We present this argument without necessarily committing ourselves to its implications; suffice it to say that it is still an open

question as to where exactly subjacency applies.

In recent years, many attempts have been made to fuse subjacency with the ECP (see section 2.8 below) to produce a unified theory of constraints on long-distance dependencies. Notable attempts include the Connectedness Theory of Kayne (see the articles collected in Kayne 1984) and developments of Pesetsky's (1982) Path Theory. We will not review these alternatives here, but instead follow the line taken by Lasnik and Saito (1984) who, while acknowledging the similarities between bounding theory and the ECP, keep the two sets of principles separate. We will, however, adopt an insight of Chomsky (1986b.) to the effect that the theory of government (as conceived of in terms of syntactic "barriers") can partially unify the two. Chomsky's version of government employs a notion of "Blocking Category" which supercedes that of bounding node:

(115) A maximal projection Γ is a Blocking Category (BC) for β iff Γ is not L-marked and Γ dominates β.

"L-marking" is defined as follows:

(116) α L-marks β iff α is a lexical category that θ-governs β.

Chomsky's notion of θ-government is extensionally equivalent to the one we have been employing; however, he treats INFL and COMP as non-lexical, which means that within his framework S

and VP are never L-marked. This necessitates an extra stipulation in the definition of "barrier" to the efffect that IP (= S) is exempt from barrier-hood. The definition is reproduced below:

(117) A maximal projection Γ is a barrier for β iff (a) or (b):

a. Γ immediately dominates δ, δ a BC for β;

b. Γ is a BC for β, Γ is not = IP.

Now, it is easy to see how subjacency can be accomodated within this model, simply by substituting "barrier" for "bounding node". However, it is rather less easy to accomodate the parametric variation associated with subjacency into the Barriers model: Chomsky proposes that the Italian/English difference noted by Rizzi might involve an extra "inherent" barrier, tensed IP (=S) in English and tensed CP (= S') in Italian; in that case, there would be an extra barrier in English but not in Italian, since CP but not IP would already inherit barrierhood from the IP which it dominates.

While many of the empirical consequences of the Barriers model are still unknown, the intuitive appeal of its "leading idea" is quite clear. We will therefore provisionally adopt this approach to the Subjacency Condition; nothing in the rest of the thesis hangs on any particular formulation.

2.8 The ECP

The central principle of "Government" theory is the Empty Category Principle (ECP). Intuitively, the ECP is a constraint on recoverability; it ensures that a (non-pronominal) empty category is in a position where its lexical content can be identified. There appear to be two main ways in which this recovery operation can take place. The first is through lexical government. Lexical governors are generally taken to be X^0 categories; for our purposes we will assume that any argument within the (internal) Case-government domain of a proper governor X is lexically governed by X. Thus a WH-phrase extracted from the direct object position of a verb will leave behind a properly governed empty category:

(118) Who_i did you see e_i ?

This approach to lexical government is discussed and rejected by Chomsky (1986b.), mainly on the basis of certain unexpectedly ungrammatical infinitival complements and the "super-raising" construction (see Chomsky, op.cit, p. 18); our (non-clausal) approach to infinitival complements (which therefore do not contain barriers, given our interpretation) renders these objections harmless.

Note that while lexical government is dependent on (internal) Case-government, it is not dependent on actual Case-marking. In other words, it is the configurational relation of minimal Case-government which allows lexical government, not Case-marking itself. Thus the empty category in raising

constructions will be lexically governed, though raising predicates do not Case-mark the subjects of their complements:

(119)a. John$_i$ is likely e_i to leave.

b. John$_i$ appears e_i to have left.

Note that raising in NP is ungrammatical:

(120)a. *John's$_i$ likelihood e_i to leave.

b. *John's$_i$ appearance e_i to have left.

This contrast has been taken as evidence that adjectives and verbs are lexical governors, while nouns are not. The ungrammaticality of extraction from a non-subcategorized PP has likewise been taken to indicate that in the unmarked case P is not a proper governor either:

(121) ??What$_i$ did you go for a walk in e_i ?

Kayne (1984) accounts for the grammaticality of most "preposition-stranding" configurations in English as well as for their cross-linguistic rarity by adapting a reanalysis account originally formulated by Hornstein & Weinberg (1981). According to Kayne, it is a marked property of English that allows verbs to optionally transfer their proper-government capability to prepositions which they properly govern. This will account for (122) below:

(122)a. What$_i$ did you talk about e_i ?

b. Who$_i$ did you talk to e_i ?

Let us assume, then, that only [+V] categories are inherently lexical governors.

The other way to satisfy the ECP is via antecedent government. Structures such as those in (123) below cannot be licensed by lexical government since the relevant empty category is not in a position where it is in the government domain of a [+V] governor:

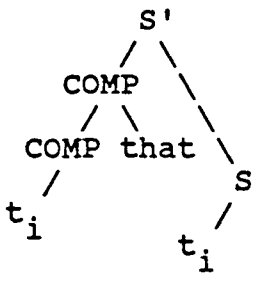
(123)a. *Who$_i$ did you say that e_i had already left ?

b. Who$_i$ did you say e_i had already left ?

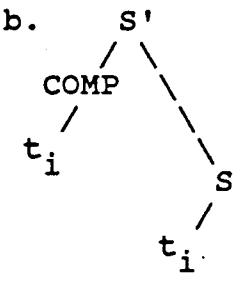
The contrast illustrated above is the classical "complementizer-trace" effect; extraction from subject position of a tensed finite clause is grammatical just as long as no overt complementizer is present. The most prevalent analysis of this contrast assumes that a trace occupies the COMP position of (123b.), from where it can antecedent govern the trace in subject position. When a lexical complementizer is present, however, antecedent government is blocked. Within a GB theory which assumes only one COMP position - as opposed to the more elaborated structure involving a "SPEC C" position" - two ways have been proposed to ensure this. The first is to assume that an antecedent must (minimally) c-command its trace, and that a

WH-phrase moving into a COMP already occupied by a lexical complementizer must adjoin to COMP, creating a branching structure where the trace in subject position cannot be minimally c-commanded. This is illustrated in (124) below:

(124)a.



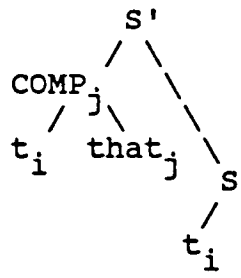
b.



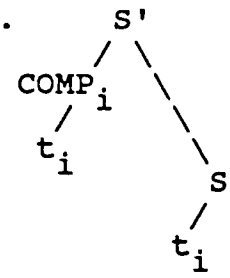
In (124a.) the trace t_i in COMP minimally c-commands only the complementizer, and not the lower trace: therefore antecedent government cannot hold. On the other hand, in (124b.) the trace in COMP minimally c-commands the trace in subject position, licensing it via antecedent government.

The second hypothesis, proposed by Aoun, Hornstein & Sportiche (1981), employs a COMP-indexing mechanism to derive the same effect. The idea is that the index of an element in COMP percolates to the COMP node itself, and that only an appropriately indexed COMP can antecedent-govern an empty NP within its c-command domain. Where an overt complementizer is present, its index will percolate to COMP and prevent the trace of a subsequently moved WH-phrase from itself indexing COMP and thereby licensing a trace in subject position. This is illustrated in (125) below:

(125)a.



b.



The more richly articulated structure that we have proposed for COMP gives us a third possibility. Let us say that antecedent-government is normally only possible from the [C] (complementizer) position. In that case, on the assumption that WH-phrases move into a separate ("SPEC C") position, antecedent-government will normally only be possible if the [C] position can bear the index of the WH-phrase. The conditions under which such indexing may apply and the parametric variation associated with COMP-indexing have been the subject of much enquiry: see Koopman (1983), Travis (1984).

As mentioned in 2.7 above, the ECP has been recently recast (by Chomsky 1986b.) in terms of a version of government based on the idea of syntactic barriers. While I do not wish to go into the complexities of this analysis at this point, it will be helpful to give Chomsky's definition of government, which reads as follows:

(126) α governs β iff α m-commands β and there is no Γ, Γ a barrier for β, such that Γ excludes β.

"m-commands" refers in our terms to maximal c-command; we may

replace it with minimal c-command, since we do not wish, as Chomsky does, to produce a single definition of government to cover all government relations. "Excludes" is relevant only for the theory of adjunction, which does not concern us here; for our purposes, Γ may be thought of as simply intervening between α and β.

Now, since the ECP is a constraint on representations, the proper government requirement may be seen as imposing a particular structure on chains of coindexed categories, so that intervening (maximal) projections do not constitute barriers to government. This is indeed the approach to antecedent government adopted by Chomsky, who in turn builds on the work of Lasnik and Saito (1984).

The ECP is generally taken to apply at LF. The evidence for this comes from the rule of Quantifier Raising (QR) which itself applies at LF (since it has no phonological effects), adjoining a quantifier phrase to S (or VP; see Williams 1983, May 1985), from where it will c-command a variable within its scope. QR appears to show ECP effects, as evidenced by the behaviour of negative quantifiers in French (Kayne (1984), and Italian (Rizzi 1982). In (127a.) below, QR has extracted <u>personne</u>, adjoining it to the matrix S, where it is interpreted as a wide-scope quantifier (as in the LF representation (127b.). QR leaves a trace which will be properly (lexically) governed by the verb <u>arrêtent</u>, thus satisfying the ECP. In (127c.), however, QR has extracted <u>personne</u> from subject position, where no lexical governor is available. Antecedent-

government will also be impossible from this position since the extracted quantifier is adjoined to the matrix S; QR is an "unbounded" movement (i.e. not subject to subjacency) and consequently leaves no intermediate trace in the embedded COMP which could "link" the trace in subject position to its antecedent: see the LF representation (127d.):

(127)a. Je n'ai exigé qu'ils arrêtent personne.
"I neg have required that they arrest nobody."

b. [personne$_i$[j'ai exigé[qu'ils arrêtent e$_i$]]]

c. *Je n'ai exigé que personne soit arrêté.
"I neg have required that nobody be arrested."

d. [personne$_i$[j'ai exigé[que e$_i$ soit arrêté]]]

Note that we are assuming that only an element in COMP is accessible to government from outside C"; thus the embedded C" node in (127d.) blocks antecedent government of the embedded trace.

Another source of evidence for the application of the ECP at LF comes from an ECP analysis of the "Superiority Condition" (Chomsky 1976). This condition accounts for the grammaticality contrast below:

(128)a. Who saw what ?

b. *What did who see ?

The Superiority Condition states (informally) that in multiple WH-constructions of this type, the "highest" WH-phrase in the

tree must move to COMP first. Chomsky (1981) derives this generalization from the ECP in the following manner: (i) any WH-phrase moved from subject position must antecedent-govern its trace (since the trace lacks a lexical governor); (ii) WH-phrases (excluding those in echo questions, which appear to behave quite differently) must move to COMP, either in the syntax or in LF (see Huang 1982, Saito 1985); (iii) a WH-phrase already in COMP will prevent a subsequently moved WH-phrase from antecedent-governing its trace (by one or other of the mechanisms already outlined above). Now, in (128a.), who will antecedent govern its trace, and subsequent LF movement of what will not violate the ECP either, since its trace is lexically governed. However, in (128b.), who will fail to antecedent-govern its trace, since COMP is already occupied by what, and the ECP will be violated, leading to ungrammaticality. The superiority effect will then follow from the ECP in just the same way as the complementizer-trace effect discussed above.

The ECP has been the focus of intense scrutiny within the GB model. Much effort has been devoted either to replacing it (Pesetsky 1982, Bouchard 1984, Aoun 1985) or to generalizing it (Kayne 1984). Much of the attention it has received has been due to its rather inelegant formulation, which incorporates a disjunction between lexical and antecedent government. Attempts to eliminate this disjunction have involved both treating antecedent-government as a special case of lexical government (Stowell 1981) and treating lexical government as a special case of antecedent-government (Kayne 1984). We will not

attempt to give an account of the complex debate that has ensued.

2.9 Learnability, Epistemological Priority, and the GREPP

2.90 Introduction

At the beginning of this chapter, we set ourselves the task of reconstructing GB theory so as to produce a syntactic model which was not only learnable (in the sense of the "logical" theory of language acquisition) but met the additional criterion of epistemological priority. In this section, we will attempt to draw together the various threads of our analysis, with a view to constructing a plausible hypothesis for the course of language acquisition. The section will consist of four parts. In the first, we will present the central hypothesis of this chapter, to the effect that a relativized notion of government can act as the link between the epistemologically primitive subcomponents of Case- and θ-theory and the epistemologically derived component of phrase-structure. In the second, we show how this relativized conception imposes a logical structure on the (real-time) sequence of language acquisition. In the third, we will compare the approach taken here with other accounts of the acquisition of phrase-structure, notably those of Pinker (1984) and O'Grady (1987). Finally, we will re-examine the role of the GREPP (our generalized version of the Projection Principle) in the context

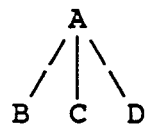
of the theory presented here.

2.91 Government and Epistemological Priority

In section 2.1, we pointed out that the central structural relation of government occupies a somewhat paradoxical position within GB theory, since though potentially unifying, no satisfactory unified version has yet emerged. We have argued throughout this chapter that, rather than being a problem for the theory, this is exactly what we should expect. We will now show why.

First of all, let us review the different types of government which we have proposed. The most restricted form of all, θ-government, characterizes the assignment of internal θ-roles, and takes place only between sisters, as illustrated in (129):

(129)



B,C, and D are all in a relation of θ-government. Note that we are also assuming a general "head-percolation" convention, so that if, for example B θ-marks C, C a maximal projection, then B will also θ-mark the lexical head of C.

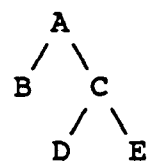
A less restricted version of minimal government characterizes assignment of "internal" Case, which can penetrate certain types of projection (those which do not

constitute barriers, in the sense of Chomsky 1986b., as discussed in 2.7 above). We have also claimed (in 2.8) that this type of government is the relevant one for the "lexical government" portion of the ECP.

Case is also subject to an adjacency condition stated on strings rather than trees, which further constrains its applicability.

"Internal Case-government" will then take place in (129) as well as in configurations such as that in (130):

(130)

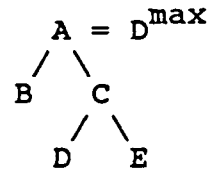


Here (assuming strict adjacency) if B is a Case-marker and C is not a barrier, B will Case-mark D but not E.

Note that we are also assuming that both internal Case and θ-marking are directional (rightward in VSO and "typical" SVO languages, leftward in SOV languages).

A second type of government relation characterizes the assignment of "external" Case and θ-roles (the latter loosely subsumed under the relation of Predication, which in turn involves the definition of AGR-domain in (43), which is based on maximal government; see 2.33). Here maximal c-command is the relevant structural relation, as exemplified in (131):

(131)



The relevant governor this time is D, which maximally governs B and minimally governs E. The order of B and C is subject to variation: as we shall see, there are languages such as German in which B and C occur in the opposite order. Note that maximal projections constitute "absolute" barriers to "upward" government; as mentioned in 2.4, there is no upward exceptional Case-marking. This imposes a further asymmetry between maximal and minimal government.

Turning to the subsystems of binding, bounding, and the "antecedent government" portion of the ECP, we see that in each case the relevant notion of government includes an extended type of minimal c-command. Thus the binding principles make reference to a (minimally) "c-commanding antecedent", bounding theory - if construed representationally rather than derivationally - involves the same relation, and the antecedent-government part of the ECP also involves minimal c-command. The differences between the locality conditions imposed on syntactic representations by these subcomponents derive from other constraints on the antecedent-anaphor relationship. Thus the notion of Governing Category is unique to binding theory, and while subjacency and the ECP both make reference to the concept of syntactic barriers, they do so in different ways: subjacency imposes constraints on how many

barriers may be crossed during a derivation (or alternatively, on how many barriers may stand between an antecedent and its trace), whereas antecedent-government superimposes a particular structure involving chain links on a version of government which is itself based on barrierhood (see 2.8 above and Chomsky 1986b., section 5).

What we see, then, is not a unified conception of government, but an array of government relations. Now, factoring out for the moment relations of linear precedence, let us construe such government relations as locality conditions on dominance - ranging from the strictly local relation exemplified by sisterhood (as in θ-government) to the "long range" locality conditions exemplified by subjacency. I wish to propose two connected hypotheses concerning the relation of such conditions to language acquisition. Firstly, I propose that first language-learners employ a relativized version of the notion of Canonical Government Configuration to map Case- and θ-relations onto tree structures; secondly, I claim that the course of first language acquisition reflects a progression from strict types of locality through to less strict long-range conditions, and that this progression reflects a corresponding "grammaticalization" of syntactic acquisition, as the child moves from dependence on epistemologically primitive subcomponents to more "grammatically sophisticated" types of acquisition. (In this sense, such a model meets the "developmental desideratum" argued for by Lebeaux 1987). It is to the first of these

hypotheses that we now turn; the rest of this sub-section will consist of a hypothetical account of the acquisition of phrase-structure configurations. In the next sub-section we will show that this hypothetical account shows promise as a model for real-time language acquisition.

We have identified θ-relations as plausible "epistemological primitives" for the language learner. This is relatively uncontroversial; we assume, along with virtually all serious work in language acquisition, that a child has prelinguistic access to information on which θ-roles go with which predicates. It should perhaps then come as no surprise that the first structural mapping performed by our hypothetical language-learner is mediated by the following θ-associated Canonical Government Configurations:

(132) θ-based Canonical Government Configurations:

I. W and Z are in an Internal θ-Government Configuration iff W assigns an internal θ-role to Z. In that case, there is a Y such that Y minimally dominates W and Z.

II. W and Z are in an External θ-Government Configuration iff Z is a predicate and W is the subject of Z. In that case, there is a Y such that Y minimally dominates W and maximally dominates Z.

Several comments are in order. First of all, note that we have

dropped all reference to linear ordering; CGC is defined in strictly heirarchical terms. Second, we have extended the notion of CGC to include Predication, which has a different structural realization than that which characterizes the assignment of internal θ-roles. Third, we have (temporarily) dropped reference to Case-marking, which will be dealt with by a separate set of CGCs. Finally, we distinguish minimal domination (equivalent to immediate domination) from maximal domination (equivalent to a version of immediate domination which only makes reference to maximal projections).

In order to identify θ-associated CGCs, the child must already know some basic categorial information (i.e., s/he must be able to identify at least some (canonical) verbs and nouns). S/he must also be able to distinguish between internal and external θ-roles. The latter task involves the setting of one important parameter - that associated with "deep" ergativity.

The "deep" ergative parameter concerns the canonical assignment of θ-roles (Marantz 1984, B. Levin 1983). In a "deep" nominative-accusative language, such as English and apparently the vast majority of other languages, the agent of a transitive verb is realized as the external θ-role, and the patient/theme as an internal θ-role; in "deep" ergative-absolutive languages (e.g., Greenlandic Eskimo) the converse holds. There is a clear markedness relation holding between the two values of this parameter, with the nominative-accusative value the unmarked option. We assume that the child learning Greenlandic Eskimo will have to encounter relevant positive evidence before

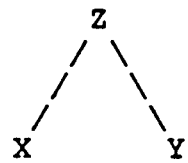
setting the parameter for the ergative-absolutive value; such evidence will be available in the child's input.

Note also that in cases where only one θ-role is syntactically realized, the argument bearing that θ-role must be externalized (in main clauses), in order to satisfy Predication. However, we assume that the child is able to distinguish between "intransitives", which have a single external argument and "unaccusatives" which have a single (externalized) internal argument. The distinction is syntactically important in many languages; for example, it determines the choice of auxiliary in French and Italian (see Burzio 1981), and it has been suggested that in English the pleonastic <u>there</u> is only inserted in the subject position of unaccusative verbs. Within a framework such as ours which retains NP traces, the distinction will be signalled by the presence of an internal argument trace in VPs containing the latter, but not the former. This in turn means a slightly more sophisticated algorithm for mapping θ-roles into argument positions must be available to the child learning a "deep" nominative accusative language: in particular, the "theme" θ-role must always be associated with the internal argument position, and the "agent" θ-role with the external argument position, even with respect to predicates which select only one argument.

Once the child has applied θ-associated CGCs, s/he will be able to begin to construct a phrase-structure representation for θ-associated elements. At this point, we assume that the

first of our phrase-structure percolation principles will come into play:

(133) In the structure



Percolation Principle I:

> Where X θ-governs Y, the categorial features of Z will be those of X.

Together with the CGCs in (132), Principle I will enable the child to produce a partially-labelled tree for a set of categories in a θ-relation.

Next, let us turn to Case-assignment. The "epistemological priority" of Case-relations is not as obvious as that of θ-relations, particularly in a relatively uninflected language like English, where Case is not phonologically overt except in certain "relic" forms, such as the pronominal system. However, we assume the A-visibility Condition (see 2.40) as a basic (unlearnt) principle of Case theory; this means that any phonologically overt argument must be Case-visible, and the child can deduce the existence of a Case-marking relation without needing specific morphological information.

Let us then propose the following Case-associated CGCs:

(134) Case-based CGCs:

III. W and Z are in an Internal Case-Government

Configuration iff W, W a Case-assigner, assigns internal Case to Z. In that case there is a Y which minimally dominates W and dominates Z, and there is no barrier X for Z such that X excludes W.

IV. W and Z are in an External Case-Government Configuration iff W, W a Case-assigner, assigns external Case to Z. In that case there is a Y which maximally dominates W and Z.

Once again, some categorial information must have been absorbed for (134) to come into effect: in particular, the child must have learnt which categories assign Case. This is a relatively easy task with respect to the assignment of internal Case, but less so with external Case, since INFL, the external Case-assigner, is often realized only as an affix. Nevertheless, if the child assumes that Case-assigners may only assign one Case, and that every (structural) Case must be assigned by some Case-assigner, even in a simple transitive sentence it will be necessary to postulate a (non-θ-marking) external Case-assigner. Now, suppose the child also assumes the P-visibility conditions, repeated as (135) below,

(135) I. Every predicate must be contained within an AGR-domain

II. Every AGR-domain must contain a Predication Subject.

III. Every predication subject must be AGR-linked.

In that case, s/he will know that an utterance containing a predicate must contain an instantiation of AGR; AGR is a property of INFL; and by the G-visibility condition, repeated as (136) below, INFL must be associated with its canonical θ-head, V:

(136) G-heads must be θ-associated.

Given this information, the child will be able to identify the external Case-assigner as INFL.

Another prerequisite distinction, that between internal and external Case, must also be grasped in order for (134) to come into play; and once again, this relates to the setting of an "ergativity" parameter - this time, the Ergative Case-marking parameter, as discussed in 2.43.

We have formulated the Ergative Case-marking parameter as in (137):

(137) C_V must be assigned.

It is easy to see how this parameter will be set in languages with a phonologically rich Case-marking system, since the Case assigned to the subject of an intransitive sentence will be the same as the Case assigned to the object of a transitive sentence. Moreover, as discussed in 2.43, the ergative Case-marking parameter is only relevant to VSO and SOV languages, since a given Case is assigned unidirectionally, The only

situation where a child might conceivably have difficulty in setting the parameter is then a VSO or SOV language with no phonologically overt Case-marking: we predict that such languages are either highly marked or do not exist.

Once the child has set the Ergative Case-marking parameter, s/he must distinguish between external and internal Case within the target language. Now, as discussed in 2.43, assignment of C_v to the external argument of E/A languages is actually internal Case-assignment. In other words, the external Case-government configuration is simply irrelevant to "surface" E/A languages. Thus once a child has identified a language as an E/A Case-marking language, s/he will identify all Case-marking configurations as taking place under CGC III above.

This leaves N/A languages. Now, since in such languages nominative Case (C_i) is obligatorily assigned, the child will be able to identify external Case in any intransitive sentence. If Case-marking is phonologically differentiated, (as for example in German) distinguishing external Case will pose no problem in transitive sentences, either; likewise, if the direction of external Case-assignment differs from that of internal Case-assignment, as in English, the child will easily learn the internal/external distinction. This leaves N/A languages which differentiate external from internal Case neither by overt Case-marking nor by linear order. Again, these will be either VSO or SOV languages. In such cases, we will appeal to the correspondance between internal θ-marking and internal Case-marking, as expressed by Kayne's original (1984)

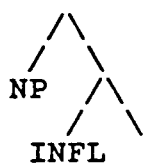
version of the CGC, which we will rename the θ-Case Correspondance Principle:

(138) The θ-Case Correspondance Principle:

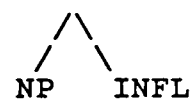
> If W and Z are in an internal θ-government configuration, and W Case-marks Z, then W and Z are in an internal Case-government configuration.

Having distinguished between internal and external Case-government, the child will be able to apply CGCs III and IV, and thereby add further structural information to the rudimentary phrase-structure representation created by CGCs I and II. Note that phrase-structure is not fully determined, however; since minimal domination is in a subset relationship with maximal domination, CGC IV allows for the possibility that an external Case-assigner minimally governs the argument to which it assigns Case, without requiring it do so. Thus in English, a child might propose either (139a.) or (139b.) for the structural relation holding between the external Case-assigner INFL and the NP to which it assigns Case:

(139)a.



b.



Once CGCs III and IV have been applied, phrase-structure

Percolation Principle II will come into play:

(140) Percolation Principle II:

In the structure (133),

Where X assigns Case to Y, the categorial features of Z will be those of X.

Thus the child is able to label the tree-structure induced by the Case-based CGCs.

As mentioned above, the Internal Case-government Configuration is identical to that required for the Lexical Government part of the ECP. The difference between Lexical Government and Case-assignment thus reduces to which categories count as Case-assigners and which count as proper governors. In the theory presented here, [V] is a Case-assigner and a proper governor, [P] a possible Case-assigner but in general not a proper governor, [I] a possible Case-assigner but again not generally a proper governor, [N] neither, [C] neither, and [N,V] a proper governor but not a Case-assigner. This is an over-simplified account of a complex set of relations, which cannot be explored further within the limited scope of this thesis.

The next step in our accquisition scenario involves G-heads and C-heads, which behave alike for the purposes of CGCs. Identifying G/C-heads should cause the child no problem, since

they assign no θ-roles and show typical "closed-class" syntactic behaviour (see Emonds 1985 for an extended justification of the open/closed class distinction within a modified GB framework). Now G/C-head association can either take place lexically (in which case the G-head will typically be realized as an affix) or syntactically. We thus need two CGCs to deal with the G/C-heads, one lexical, one syntactic. These are given in (141) below:

(141) G/C-head CGCs:

IV: Lexical G/C-head GC:

W and Z are in a lexical G/C-Government Configuration iff W a is G/C-head and W and Z are ordered in the (lexical) Canonical Government Direction. In that case, there is a Y, Y a word-level category, such that Y minimally dominates W and Z.

V: Syntactic G/C-head GC:

W and Z are in a syntactic G/C-Government Configuration iff W is a G-head and W and Z are ordered in the (syntactic) Canonical Government Direction. In that case, there is a Y such that Y minimally dominates Z and W.

It will be observed that we have now introduced directionality into the formulation of the CGCs, in the form of the concept of Canonical Government Direction, which we define as below:

(142) Syntactic Canonical Government Direction (CGD I):

> The Syntactic Canonical Government Direction of a language is the direction in which internal Case is assigned.

Again, (151) captures part of Kayne's original CGC insight; having factored out relations of linear precedence from those of dominance, we now reintroduce them separately in the form of the CGD. Note that our formulation of the CGD differs from Kayne's directional version of the CGC in making no reference to direction of θ-marking; this is to accommodate the work of Travis (1984, Chapter 2), who argues that the direction of Case- and θ-marking can differ within a single language. It may be that some categories are important as "canonical" Case-assigners and/or θ-role assigners; Travis (1984) suggests that [P] is a canonical Case-assigner, and Kayne (1984) builds the canonical relationship between a [V] and its [N] complements into his definition of CGC. We choose Case as the canonical directional relation since it is more "transparent" (in the sense of surface-recoverable) to the language learner; we speculate, however, that languages which differ in direction of Case- and θ-role assignment are in any case quite highly

marked.

Note also that the CGD applying in the lexicon may differ from that which applies in the syntax, as indeed it does in English, where the CGD is leftward inside words and rightward between words. This is not unexpected, since the syntactic CGD is decided by Case-relations, which do not apply inside words. (Note however that the "Mirror Principle" proposed by Baker 1985 makes the claim that, at least in the unmarked case, morphological and syntactic structures will be isomorphic. We will not investigate this claim here).

Word-internally, the CGD captures the difference between prefixing and suffixing languages, and may be expressed by the following informal definition:

(143) Lexical Canonical Government Direction (CGD II):

> If a language is predominantly suffixing, then the lexical CGD is leftward; if predominantly prefixing, then the lexical CGD is rightward.

Once CGCs (IV) and (V) apply, Percolation Principle (IV) comes into effect:

(144) Percolation Principle IV:

> In the structure (142),

Where X and Y are in a CGC, X governs Y in the CGD, no Case or θ-relation holds between them, and both are part of the subcategorization set of Z, the following heirarchy determines which features will percolate:

(a) C-features of X and Y will percolate to Z
(b) G-features of X and Y will percolate to Z
(c) θ-features of X and Y will percolate to Z

Note that we have altered Principle IV so as to accommodate our reformulation of the notion of CGC and our introduction of the separate concept of CGD.

The child who has reached the stage where Principle IV applies will be able to induce labelled tree structures for what we have called the subcategorization set of a given phrase. In effect, the representation of the subcategorization set is equivalent to the level of NP-structure which we introduced in 2.57.

At this point, let us pause and review the progress we have made so far. At the outset of this section, we set ourselves the task of producing an explicit account of the acquisition of phrase-structure which (a) eliminated the need for phrase-structure rules and (b) met the criterion of epistemological priority. The two requirements are linked, in that we have argued that phrase-structure rules (including X-bar schemata) do not meet the epistemological priority criterion.

We have now constructed the basis of such an account, though

many details remain to be elucidated. As always, however, it is important to calculate the costs as well as the benefits of such an analysis. In this case, the "costs" consist of the machinery we have employed - and attributed to the language learner - in order to replace the learning of phrase-structure rules. (145) is an inventory of the principles and parameters we have employed so far:

(145)a. Canonical Government Configurations (I-VI).
b. Canonical Government Directions (I-II).
c. Percolation Principles (I,II and IV).
d. The "deep" and "surface" Ergative parameters.
e. The A-visibility, P-visibility, and G-visibility conditions.
f. The θ-Case correspondance Principle.
g. The GREPP.

(145b.) can be learnt from the interaction of CGC I with Primary Linguistic Data; otherwise, with the possible exception of the GREPP (to which we will return in 2.96), the principles and parameters in (145) are part of the child's (innate) linguistic ability.

Now, it might be argued that we have merely replaced epistemologically implausible phrase-structure schemata with epistemologically implausible CGCs and percolation principles. However, the CGC/percolation-based account has a number of non-trivial conceptual advantages over more traditional phrase-

structure accounts. In particular, while a phrase-structure schemata-based model must employ additional mapping principles in order to explain how a language-learner applies PS schemata to Primary Linguistic Data (as we will see in 2.93), we have built this mapping property directly into the grammar. In this sense, the model proposed here comes closer to addressing one of the central questions of generative grammar - how a grammar is acquired, given standard assumptions about the type of evidence available to the child. Moreover, as argued throughout this chapter, a theory which meets the criterion of epistemological priority is to be preferred over one which does not.

In addition, as suggested at the beginning of this section, the model presented here offers a possible answer to certain puzzling questions within GB theory itself. As mentioned above, the concept of government has played a paradoxical role in the development of the theory, since though it enters into virtually every subcomponent, no unified version has been forthcoming. Within the theory presented here, this is precisely what is to be expected; as the central mapping relation between syntactic subcomponents, government imposes a series of interconnecting heirarchical locality conditions on phrase-structure configurations. Moreover, as we shall argue in 2.92 below, the real-time course of language acquisition shows a progression from more strictly local types of relation - those which characterize epistemologically primitive syntactic subcomponents - through to less strictly local relations, which

characterize long-distance relations. This progression is mirrored in the types of government which characterize the diferent subcomponents, providing further evidence for the importance of government not only to the theory of grammar but to the theory of language acquisition, to which we now turn.

2.92 Government and Language Acquisition

In this sub-section we will briefly examine how the predictions made by our hypothetical model of the language learner accord with what is known about the real-time course of language acquisition. We will not attempt to go into great detail; rather, our aim is to show firstly that the overall development of grammatical knowledge corresponds fairly closely to the predictions made by the model outlined above, and secondly that the kind of approach we have taken shows promise for a more detailed examination of the relation between grammatical theory and language acquisition than has hitherto been possible.

First of all, let us summarize the predictions made by our theory:

(146)a. most <-- Epistemologicallly prior --> least

θ-theory > Case-theory > G/C-theory > X-bar theory

b. most <-- Syntactically local --> least

internal Case \ binding/bounding
lexical government / antecedent government
/
internal θ - G/C-association
\
external Case > predication

(146) shows (approximately) the correspondance we have demonstrated between syntactic locality (as exemplified by government relations) and epistemological priority. "G/C-theory" is the subcomponent which deals with G-heads, as embodied by CGCs V and VI and by Percolation Principle IV. We will not discuss the "long distance" subcomponents of binding theory, the ECP, and bounding theory, partly for simplicity's sake, partly because they do not bear on later chapters.

Now, if this model is to serve as a plausible hypothesis about the real-time course of language development, we should expect a similar correspondance to show up in language acquisition data; that is, we should expect θ-relations to be the first type of grammatical information encoded by the child, followed by Case-relations, and so on. As we shall show, such a correspondance does indeed appear to exist.

As far as the priority of θ-relations is concerned, there is little controversy. In fact, at an early stage of acquisition, "semantic relations" appear to be <u>all</u> that are encoded in the syntactic representations of first language learners, as shown

in a plethora of work, including that of Bloom (1970), Brown and associates (Brown 1973), and Bowerman (1973).

The acquisitional priority of (abstract) Case relations is rather more difficult to establish. In languages with rich morphological case-systems, it does appear to be the case that the inflectional morphology associated with Case is learnt at an early stage of acquisition (Slobin 1985, Hyams 1986). On the other hand, in languages such as English which retain only relic morphologically case-marked forms, there is evidence that such forms are learnt late and with some difficulty. In English, such evidence comes from studies on the acquisition of the pronoun system (Huxley 1970). In fact, it is easy to reconcile this apparently contradictory evidence, if the assumption is made that there are two ways in which a language may (primarily) exhibit Case-relations. The first, which characterizes "inflectionally rich" languages, is morphological; the second, which characterizes "inflectionally deficient" languages, is configurational. Thus we predict the early acquisition of case-morphology in languages such as Turkish and Russian, but not in languages such as English or Chinese; on the other hand, in the latter type of language we expect Case-based word-order parameters to be set early. Precisely such a proposal has been made by Hyams (1986).

In the next two subsections we will compare the approach to the learning of phrase-structure adopted here with two others, those of Pinker (1984) and O'Grady (1987).

2.93 Pinker's Model of the Acquisition of Phrase-Structure

In this subsection we will compare our approach to the acquisition of phrase-structure configurations with that of Pinker (1984, Chapter 3). Though Pinker's syntactic assumptions are those of Lexical Functional Grammar (see Bresnan 1982), making the comparison somewhat complex, Pinker's proposals are by far the most detailed and explicit to have been made so far; as such, they set a standard for any alternative analysis.

Pinker provides six phrase structure procedures and one lexical procedure: these are listed in (Pinker, op. cit., ps. 67-68.) Some of these are not of direct interest to us, either because they must be assumed by any model, or because they concern aspects of phrase-structure acquisition which do not directly relate to the "mapping property" which primarily concerns us. Thus P.1 , for example, simply tells the learner to apply existing procedures to a newly encountered string, an uncontroversial move, while P.5 is concerned with "rule strength", and thus does not constitute a mapping procedure but a hypothesis about the "reinforcing effect" of Primary Linguistic Data. P.6 is concerned with abbreviatory conventions for re-write rules, which do not concern us because we are assuming phrase-structure rules are conditions on trees, not strings (for justification, see Gazdar et al. 1985).

This leaves P.2, P.3, P.4, and L.1. The first part of P.2 is a partial statement of the "semantic bootstrapping hypothesis", which pertains to the learning of lexical categories rather

than tree-structures. The second is more directly relevant:

(147) "...Build a branch extending each lexical category upward to its maximal projection (i.e., X" according to the version of X-bar theory adopted here). S is the maximal projection of the head of the sentence. The head of the sentence is the V" encoding tense, aspect, and modality if there is one among the major constituents in the string; otherwise it is X", where X is the major predicate of the proposition encoded in the sentence."

It will be observed that Pinker adopts many of the X-bar assumptions which we have rejected: in particular, Uniformity plays a crucial role in P.2, since it is used to define the notion of maximal projection, which in turn is crucial to P.3 and P.4. The criticisms of Uniformity made in 2.51 and 2.57 apply here; as argued in those sections, there is neither empirical evidence nor theoretical justification for such a condition, and certainly it fails to meet the criterion of epistemological priority.

Let us then turn to the first part of P.3:

(148) "Annotate the maximally projected node subtended by each argument word with the equation FUNCTION = , using the grammatical function specified for that argument by the semantics of its relation to its predicate (e.g. SUBJ for agent of action). Add the equation = to all nodes

representing heads and specifiers."

Here comparison of the mechanics of phrase-structure acquisition is difficult because of the particular LFG assumptions which Pinker employs. The assumptions themselves, however, may be examined. In particular, the role of grammatical relations in LFG (where they are primitives) has already been subject to criticism on the grounds of epistemological priority (by Chomsky 1981). In fact, Pinker does not treat grammatical relations as acquisitionally prior; instead, he uses an extension of semantic bootstrapping to map thematic roles onto grammatical relations. While this avoids some of Chomsky's criticisms, a theory which treats grammatical relations as purely derivative from the mapping of thematic relations onto configurations makes (148) entirely redundant; this constitutes a clear advantage for such a model over that employed by Pinker. Note also Pinker's (unexplicated) use of the notion of "specifier". As pointed out in 2.51 and by Pullum (1985), "specifier" is cover-term for a set of syntactic elements (including subjects and G-heads) with very different syntactic properties; it seems then implausible to attribute it primitive status within a theory of phrase-structure.

Let us turn to the second part of P.3, which deals with (morphological) Case-relations:

(149) "If the lexical head of a constituent is marked for case.., append the equation CASE = n to the phrase,

where n is the case signified by the inflected head according to the inflectional paradigm for case in the language."

The LFG concept of case is "concrete", as opposed to the abstract Case employed by GB. Thus, whereas we have treated Case as an important primitive concept (particularly in defining the syntactic CGD, which in turn enters into the determination of the CGCs for G-heads), the procedure for learning case in (149) bears little relation to the learning of other components of the LFG system. In so far as the GB version of Case-marking unites morphological case (in "inflected languages" and ordering parameters (in languages with strict word order) it attains a significant level of generalization which the LFG version misses. This has implications both for cross-linguistic studies and for the study of language acquisition. For example, Simpson (1982) notes (within an LFG framework) that in the "non-configurational" language Warlpiri (morphological) case-marking is obligatory in NP just in case elements of NP are non-contiguous; this falls out automatically if Warlpiri can realize Case either by word-order and adjacency parameters, or by morphological case-marking, as predicted by a theory of abstract Case, but this generalization cannot be directly captured in a theory such as LFG which does not link the two. In 2.92 we briefly examined the relationship between different realizations of abstract Case in language acquisition, and showed that at least on a preliminary

examination of the relevant data, children acquired morphological case early in languages with rich inflectional morphology, later in those which realize Case primarily through word-order. Again, this follows naturally if word-order and Case morphology are alternative realizations of the same abstract relation, but has to be stipulated separately in a theory where they are not.

The last part of P.3 deals with a (partial) procedure for acquiring A'-dependencies (i.e., topics). Here we have little to say; the procedure necessary to differentiate A'-positions from A-positions will be very similar for any model of phrase-structure acquisition, involving the location of a constituent in a position which cannot bear a θ-role and linking it to a position which can.

Let us then turn to P.4, which also consists of three parts. The first is as in (150):

(150) "Connect the SUBJ noun phrase as the daughter of the root S-node."

Again, (150) reflects the difference between grammatical relations-based theories such as LFG and configurationally-based theories such as GB. Since the notion of SUBJ is not a primitive within the GB model which we have adopted, there is no equivalent of (150) within our acquisition procedures. We have, however, employed the concept of "external argument", which is defined by the Predication relation together with the

assignment of external Case. Thus the work of (150) is done within our framework without ever needing to refer to grammatical relations; as argued above, this constitutes an advantage in terms of epistemological priority.

The second part of P.4 reads as follows:

(151) "Connect the remaining branches according to the information in the uncommitted f-structure and the X-bar principles (e.g., functional argument = sister of X, modifier = sister of X') and the analagous conditions for specifciers."

(151) comes closest to the notion of CGC which we have employed to map θ-relations onto tree structures. The LFG concept of "functional argument" is equivalent to our "internal argument" (note that Pinker distinguishes subjects by assigning them a separate phrase-structure procedure - i.e., (150) - which applies prior to (151)). There is no equivalent in our model to the concept of "modifier"; since modifiers form part of the adjunct set of a given phrase. As for "specifier", we have already pointed out that there is no epistemologically primitive basis to this concept and that it should therefore not be employed in phrase-structure acquisition procedures.

Note also that since in our model the Succession property of X-bar grammars is derivative, no reference can be made to intermediate X-bar projections in defining phrase-structure acquisition procedures; in other words, no mapping procedure

such as that outlined in (151) is possible within our framework. We have argued that this is an advantage insofar as Succession in a standard X-bar model such as that of Pinker depends in turn on Uniformity, which is implausible in terms of epistemological priority.

The final part of P.4 deals with cases where the procedures in (150) and (151) lead to crossing branches. Since crossing branches are impossible within a framework where phrase-structure principles apply to sequences of "local" (one-generation) trees (see 2.53 above) this situation never arises in our framework, and we can dispense with P.4 (c) altogether.

Finally, we turn to Pinker's L.1; since Pinker's language-learner (like our own) proceeds in an "analysis-by-synthesis" fashion, lexical information may be used (together with the various principles we have proposed) to induce phrase-structure, and conversely, existing phrase-structure may be used to adduce information about lexical items. This is essentially (within the specific assumptions of Pinker's model) the role of L.1. Since we do not question the need for an equivalent procedure in our own model, we will not discuss it further here.

One more aspect of Pinker's phrase-structure acquisition procedures merit discussion. This concerns the learning of "non-canonical thematic mappings" (i.e., those determined in GB by NP-movement rules and in grammatical relations-based frameworks such as LFG by function-changing rules).

Pinker (op.cit., ps. 76-83) discusses the acquisition of

passive structures in some depth. Since in LFG c(onstituent) structures are treated separately from f(unctional structures, there are two parts to Pinker's analysis: the first deals with learning the phrase-structure configurations associated with passives, the second with learning the correct θ-role assignment. The child accomplishes the first task in Pinker's model by noting that oblique passive by-phrases have the same distribution and internal structure as adjunct PPs; the second task is accomplished by adding a new lexical entry, appropriately annotated with functional equations, to the child's lexicon.

Some of the assumptions in Pinker's analysis must, I think, be made in any account. In particular, the child must recognize that, at least with respect to "reversible" passives, active sentences are in some sense "basic" (or epistemologically primitive). This is reflected in the child's input; Pinker reports an unpublished study in which he found that reversible passives with oblique by-phrases are uncommon in speech to young children. This assumption is particularly important in an analysis such as the one we have proposed where certain types of θ-role are (subject to the Ergative θ-marking parameter) intrinsically associated with the "external argument" position.

The present analysis differs from Pinker's in that the GREPP ensures the presence of a trace in the object position of passivized sentences. Thus there is no θ-role associated with subject position; the subject NP is θ-marked via coindexation with the internal θ-position, just as we have proposed for

"unaccusative" verbs.

This entails a difference in the lexical entries for passivized and non-passivized verbs from that proposed by Pinker. Pinker's version is as in (152):

(152)a. bite: V: PRED = "bite (SUBJ, OBJ)"

b. bitten: V: PRED = "bite (OBL_{agent}, SUBJ)"

In our framework there is no need to specify that the single θ-role of a passivized verb appears in subject position; passivization simply involves the suppression of the agent θ-role:

(153)

Recalling the discussion of predicative <u>be</u> in 2.55, the empty external θ-slot associated with <u>be</u> links to the internal argument of <u>bite</u>, which forces "NP-movement", or rather the formation of a chain linking the two positions.

As far as the child is concerned, then, we should expect "truncated" passives to be as easy to learn as unaccusatives. The literature bears this prediction out; as is well known, non-reversible passives appear long before reversible passives (see Horgan 1978, Maratsos 1978). We also predict that the <u>by</u>-phrase associated with oblique agents should be learnt quite

separately from the actual passivization process, a prediction which appears quite plausible in developmental terms: see the discussion in Pinker, ps. 80-81, and references cited there. Note that since we do not mention the oblique agent in the lexical rule (153), it will be realized (if at all) as an adjunct; this accounts at once for its distribution (that of any other adjunct PP) and its optionality. We also avoid the problems associated with Pinker's account of the acquisition of passive by, where the child initially attaches the by-phrase as a sister to [V], and then is forced to retreat when faced with sentences such as (154):

(154) The dog was bitten in the park by a cat.

Here the oblique agent is attached outside an adjunct PP, and therefore cannot be a sister to [V] without crossing lines.

This account is not intended to explain all the intricacies of the acquisition of passive structures; our aim has been merely to show that passive can be acquired under our theoretical assumptions, and that our model does at least a good a job of it as that proposed by Pinker.

To conclude this subsection, we point out the two main advantages of our account over Pinker's. To be fair to Pinker, the deficiencies in in his phrase-structure procedures are those of the grammatical model he employs, not of his approach to language acquisition, to which we are fully sympathetic.

The first concerns Pinker's use of X-bar theory, which we

have argued is implausible from the point of view of epistemological priority. In particular, Pinker's version of X-bar theory is crucially dependent on Uniformity, which we have argued is both theoretically and empirically flawed, and he is forced to make use of notions such as "modifier" and "specifier" whose only definition is in terms of the very phrase-structure configurations which they are meant to define.

The second major difference is in the status of grammatical relations, which as originally argued by Chomsky (1981) cannot be taken as primitives within a theory of language acquisition. By eliminating them, we have effectively removed a whole level of mapping from the language acquisition procedure, which Pinker is forced to retain.

We conclude that the model presented here has certain conceptual advantages over that proposed by Pinker. Again, it should be emphasized that these advantages are those of the grammatical theory which we employ, not the way in which we apply it to language acquisition. This is exactly as it should be, if, as we have argued, a grammatical theory at a certain level of explanatory adequacy must meet the criteria set by language acquisition; only by meeting such criteria do we begin to approach the central questions of Universal Grammar.

2.94 O'Grady's Model of the Acquisition of Phrase-Structure

William O'Grady (1987) produces an account of the acquisition of phrase-structure within the context of a more general claim

that "descriptively adequate grammars can be constructed from concepts and relations that are not specific to the language faculty and that the ontogeny of these grammars does not presuppose the existence of innate linguistic knowledge." (O'Grady 1987, p.1). This is referred to as the hypothesis of "general nativism" as opposed to "special nativism".

While it is not my intention to enter into the general/special nativism debate in any detail, I think it can be shown that the particular phrase-structure model which O'Grady adopts (one which in fact shares many salient characteristics with the one proposed here) in fact provides arguments for special rather than general nativism. Let us then turn to the details of his model.

O'Grady adopts a modified categorial grammar which is similar to that proposed in this chapter in at least two significant ways: it is a "projection" of the lexicon, in that phrasal categories are projections of lexical categories; and it employs principles of categorial combination which are based on a notion of syntactic "dependency" very similar to the ideas of categorial association and saturation adopted here, and probably even closer to the model of Higginbotham (1986), which employs a more unified notion of saturation. In fact, both Higginbotham's model and that of O'Grady share a certain confusion over whether saturation/dependency is essentially a semantically-based notion - in which case, their models would essentially turn out to be types of Montague grammar, employing a rule-to-rule mapping approach (see Bach 1976) - or a

syntactically-based notion, in which case their approaches would more closely resemble that of Speas (1986) and myself. O'Grady in fact claims that his position is closer to the syntactically-based version, though at crucial points he appeals to an (unelucidated) syntax-semantics correspondance.

In addition, O'Grady assumes the following principles:

(155) The Adjacency Principle (AP):

Combine adjacent elements.

(156) The Heirarchical Structuring Requirement (HSR):

Combinatorial operations create phrases to which subsequent combinatorial operations apply.

(157) The Subject-Last Requirement (SLR):

The subject combines with VP (the maximal verbal category).

(158) The Phrasal Category Principle (PCP):

A phrase takes on the category of a component with an unsatisfied dependency.

Most of these principles have more or less direct analogues in the model adopted here: The AP is partially subsumed under Case-theory; the HSR is simply the principle of Structure-dependency (see Chomsky 1975); the SLR is accounted for by principles of external Case-assignment and Predication; and the PCP is a version of phrase-structure saturation.

Now, so far nothing special distinguishes O'Grady's model from the others discussed here. However, it is O'Grady's contention that all of the principles above are derivative from the interaction of more general learning mechanisms with the

PLD; this is of course the hypothesis of general nativism. In order to assess this claim, let us look in slightly more detail at the relevant arguments.

First of all, O'Grady notes that the basic relations of dependency and adjacency are common to all cognitive systems, and thus plausibly available to a general learning mechanism.

Second, O'Grady assumes (along with virtually all other work on language acquisition) that learning is a mapping from a pre-existing "semantic representation" onto a linguistic representation, and that this "semantic representation" is non-linguistic.

Third, O'Grady assumes a general rule learning mechanism - "generalize", constrained by two other requirements, the Conservatism Thesis reproduced as (159) and the Trigger requirement in (160):

(159) The Conservatism Thesis:

> Children make use of the available concepts to formulate the most conservative hypothesis consistent with experience.

(160) The Trigger Requirement:

> No change is made in the grammar without a triggering stimulus in the environment.

In order to avoid the "generalization" hypothesis contadicting the Conservatism hypothesis, O'Grady claims that the latter applies to the formulation of principles, the former to their application.

Given these assumptions, O'Grady's claim is that the

syntactic principles (155-158) are derivable from the interaction of general cognitive principles with PLD. Let us take a concrete example. In a sentence such as

(161) Harry left Mary.

the child must (a) learn that there is heirarchical structure underlying the string (i.e, the HSR); (b) that Harry and Mary are independent categories and left a function which maps them into a proposition, rather than the other way round; (c) that left Mary rather than Hary left or Harry Mary counts as an "intermediate projection" (by the AP and the SLR); (d) that V is the head of the VP rather than, say, N (by the PCP).

O'Grady's position is that all this is learnable after "exposure to even a small sampling of simple syntactic structure/semantic form mappings" (p.28), given the availability of adjacency and dependency. In fact, the key to his approach is what I shall refer to as the "syntax/semantics isomorphism"; essentially the child simply structures syntactic categories in exactly the same way as the corresponding semantic representations, unless evidence is encountered to the contrary. Thus, since certain semantic representations are structure-dependent, so will the corresponding syntactic representations be, unless evidence is encountered to the contrary; since the child naturally generalizes, s/he will then hypothesize that all syntactic structures are subject to the HSR, accounting for (a) above. Similarly, since nominals are

semantically referring expressions denoting entities, but verbs are functions from entities to truth values, the child will correctly structure the equivalent syntactic dependencies, accounting for (b). The AP will be hypothesized on the basis of the fact that in general elements in a dependency relation are adjacent, partially accounting for (c). The SLR, which accounts for the other half, is acquired on the basis of the AP together with the child's tendency to generalize; thus a child who has learnt the AP will attempt to apply it generally, and will prefer to hypothesize a continuous than a discontinuous constituent for structures such as (162), where the adverb is (semantically) a VP-modifier:

(162) John runs fast.

This will provide evidence for a VP constituent, and thus for the SLR. Finally, (d), which derives from the PCP, is again straightforwardly derived from the syntax-semantics isomorphism, in that just as a a semantically dependent category combimes with a semantically independent category to yield another semantically dependent category, so a syntactically dependent category will combine with a syntactically independent category to yield another syntactically dependent category.

There are essentially three questions to be asked about this type of scenario. First, does it work ? Second, does it capture the right type of generalization ? And third, does it do all

that its author claims ?

The answer to the first question, though unknown until someone actually tries to run it through a computer, appears to be affirmative; as a discovery procedure, there seems no reason to doubt the mechanical adequacy of O'Grady's model.

The second question is inherently more interesting, and it is considerably less clear that O'Grady's model makes the right type of generalization. In particular, note that for O'Grady, the entire syntactic component is essentially derivative from the interaction of learning principles and the semantic representation. Thus any purely syntactic generalization - either cross-linguistic or acquisitional - will be hard to capture. But of course, such generalizations are numerous: the whole idea of an autonomous syntactic component in natural language springs from the failure of semantic representations to plausibly capture them. Naturally, it could be that we have simply not yet got the right semantic representation; but the language-particular nature of such a representation would then in any case undermine O'Grady's theory of general nativism, since semantic representations could not be taken as "general cognitive representations", as O'Grady claims, following Jackendoff (1983).

This brings us to the third question. Does O'Grady's learning procedure support the hypothesis of general nativism ? The answer would appear to be no, as long as his theory is intended to capture the type of syntactic generalization alluded to above; either semantic representations must be

language-specific, or there must exist language-specific syntactic principles, or both.

Another claim that does not appear to be borne is that O'Grady's learning procedure is distinct both from the "semantic bootstrapping" approach of MacNamara (1982) and Pinker (1984) and the "distributional learning" model of Maratsos and Chalkley (1980). In the former type of learning, there is a "canonical structural realization" (see Grimshaw 1981) between semantic entities (e.g., "things", "attributes", etc.) and syntactic categories (N, V, etc.), which enables the child to initially construct "semantically modelled" syntactic structures - after which, purely syntactic types of learning come into play. In the latter type, learning is purely distributional; the child creates form classes based on syntactic regularities, in much the same way as a structuralist linguist, independently of semantic representations.

In fact, it appears to me that O'Grady's model (probably, like any other even vaguely sensible proposal) employs a combination of bootstrapping and distributional learning. There must be bootstrapping in order to "enter the system", since the concept of "independent category" necessary for the semantics/syntax isomorphism to operate is semantically rooted in the world of referents (i.e., things) and propositions (i.e., statements about things). There must be distributional learning in order to acquire the AP, the SLR, and so forth. In other words, O'Grady's hypothesis is not an alternative to either semantic bootstrapping or distributional learning, but a

particular combination of them.

To conclude: while O'Grady presents a reasonably plausible learning mechanism (actually, one that might be adopted by Montague grammarians, for example, since it involves an acquyisitional version of the rule-to-rule mapping hypothesis, in the form of the syntax-semantics isomorphism), the mechanism affords no support whatsoever for the claims of general nativism, since it must either presuppose a linguistically specialized semantic component, or a linguistically specialized syntactic component, or both.

This concludes our review of alternative phrase-structure acquisition mechanisms. I hope to have shown that the particular version I have adopted (a) constitutes a plausible model for the learning of phrase-structure within a GB framework; (b) meets the criterion of epistemological priority set out at the beginning of the chapter; and (c) is consonant with what is known about real-time language acquisition.

We now turn to a particular issue in the acquisition of syntax, where I argue that the derivative nature of phrase-structure in the model I have proposed provides a simple solution for a set of data which has generated a considerable amount of recent controversy in the language acquisition literature: the so-called "missing subject" stage in English.

2.95 Subjectless Sentences and the Pro-Drop Parameter

In this section, we will turn to a particular "case-study"

which has recently attracted much attention in the field of language acquisition; the "missing subject" stage in early language acquisition. I will argue that, contrary to recent analyses, including those of Hyams (1983), Guilfoyle (1984), and Lebeaux (1987), the missing subject stage presents no evidence for or against a parameter-setting model; instead, it represents a "presyntactic" period of linguistic development, in the sense that at the period in question, children have not yet incorporated Case- and θ-relations into fully-articulated phrase-structure representations.

Non-imperative "subjectless" sentences are common in the corpora of Bloom (1970), Braine (1976), Brown (1973), and Bowerman (1973), and more generally in the language of children up to approximately three years old (or, of MLU up to 2.75), whereas "objectless" sentences are not. Two main types of explanation have been proposed for this: the "complexity hypothesis" of Bloom (1970), and the "grammar-based" hypotheses of Hyams (1983), Guilfoyle (1984), and Lebeaux (1987). We will briefly review these explanations, showing that our alternative avoids some of the dubious assumptions necessary in both accounts.

For Bloom, the occurrence of subjectless sentences is a surface phenomenon of syntactic "reduction". She operates under the assumption that fully specified phrase-markers underly such sentences, but that performance constraints force the child to reduce his/her utterance by dropping subjects. However, as Hyams (1983, p.224 ,n.8) points out, this does not explain why

it should be subjects that are consistently dropped, nor why there is no precise correlation between "complexity" and the lkikelihood of reduction. More generally, since Bloom does not address the issue of how "performance constraints" apply, her analysis in effect merely transfers the question of subjectless sentences from the theory of the acquisition of syntax to a hypothetical theory of the application of performance constraints to syntactic acquisition, a move which we crticized in Chapter 1.

The analysis of Hyams (1983) is at the opposite extreme. Hyams hypothesizes that children learning English have in effect set the pro-drop parameter the wrong way, mistaking English for a language which, like Italian, employs the phonologically empty "little pro" in subject position. Now, since there is abundant evidence for the language-learner that English is <u>not</u> a pro-drop language (lexical subjects, including expletives, are obligatory), Hyams is forced to assume that the pro-drop parameter is set by the child prior to encountering any evidence. This means that the child learning any non-pro-drop language will be forced to retreat from his or her first, prelinguistic hypothesis, on the basis of the evidence in the Primary Linguistic Data (i.e., lexical expletives).

There are several problems with Hyams' analysis, some of which have been touched on in Chapter 1. We will focus on three: the first concerns the problem of "triggering data", and is not specific to Hyams, but constitutes a more general issue

in the parameter-setting approach to language acquisition; the second is syntactic, and concerns the pro-drop parameter itself; and the third is empirical, and concerns the data on which Hyams bases her analysis.

The problem of triggering evidence is the following: given that English shows no pro-drop characteristics - in fact, much evidence to the contrary - why should the child "ignore" (or filter out) the relevant evidence up until a certain point in acquisition, when it suddenly becomes salient - in fact, vital - as the trigger for parameter-resetting ? Hyams herself recognizes this problem ,which she calls "the problem of selective attention", without proposing a solution. Note that no appeal can be made to the inaccessibility of the evidence itself, since sentences as simple as those in (163) should be able to act as pro-drop triggers:

(163)a. It's time for bed !

b. It's snowing outside.

c. There's a fly in my soup.

d. Once upon a time, there were three bears...

The second problem we will discuss is more strictly syntactic in nature, and necessitates a brief discussion of Hyams' treatment of the "pro-drop" parameter.

For Hyams, pro-drop involves PRO (i.e., the ungoverned pronominal anaphor) in AGR. The features of PRO will in turn "identify" pro, the governed empty pronominal, in subject

position (see Rizzi 1986, Roberge 1986 on "identification"). One of the consequences of this analysis is that pro-drop is incompatible with lexical material in INFL (since such material would govern PRO), therefore any language which allows auxiliary elements in INFL (such as English or French) should disallow pro-drop.

Now, while ingenious and initially plausible, this analysis has some serious flaws; in particular, it conflates two quite <u>different</u> types of "null argument". As pointed out by Huang (1982) languages which allow null arguments are of two different types: they either have "rich" AGR or no AGR at all. The former type is exemplified by Italian, Spanish, Irish, and Hebrew; the latter by Chinese and Japanese. The two types are differentiated by a number of criteria; for example, the phenomenon of "paradigm-linked" pro-drop (see Borer 1984, McCloskey and Hale 1984) in which only certain person and number inflections are "rich" enough to allow null subjects, is characteristic of AGR-linked pro-drop, while the "0-topic" phenomenon (see Huang 1982, Leijong 1986) in which null arguments appear to be discourse-bound rather than identified by syntactic features is characteristic of non-AGR-linked pro-drop.

Now, the point of all this is that Italian is certainly of the AGR-linked type; and as shown by Rizzi (1982, 1986), Jaeggli (1982), Borer (1984), Roberge (1986) and much other work connected with the recoverability of null subjects, the most important condition for AGR-linked pro-drop is the

existence of a "rich" AGR(eement) system, which can "recover" the syntactic features of the empty pronominal pro. However, at the "null-subject" stage in Hyams' analysis, English children clearly have not mastered the AGR-system at all; thus if they have set a "null-subject" parameter the wrong way, it is in the direction of Chinese, not Italian. Now, the latter type of parameter setting may be plausible; but it is not the one assumed by Hyams.

Next, let us turn from theoretical to empirical considerations. There are three purely empirical problems with Hyams' analysis. First of all, at the null-subject stage - which lasts until the acquisition of the auxiliary system - Hyams assumes that subjects are optional in English, as in adult Italian. Fair enough; this has been observed by many authors, some of them cited above. However, she neglects to mention that at the same stage, declarative sentences consisting of two constituents but missing a verb are as or more common than subjectless sentences. This is clearly documented in Wells' (1985) large-scale study (see Table A18), where both the occurence of subjectless and verbless declarative sentences rises to a peak in the 24-30 month age-range and thereafter declines at approximately the same rate. Since no version of the pro-drop parameter will account for verbless structures, we are justified in feeling somewhat suspicious of Hyams' parameter-setting account - or indeed, any purely grammatically-based explanation.

Second, Hyams' is committed to the view that at the null-

subject stage, English children have not yet properly analyzed auxiliaries. In fact, in the period in question (corresponding to Bellugi's (1967) period B, or MLU stage III) it is frequent for children to have acquired one or two auxiliaries (typically, can't or don't along with the "semi-auxiliaries" gonna and hafta. Hyams is forced by her assumptions to treat the former as allomorphs of not, the latter as (properly inflected) verbs. There at least three problems with this very ad-hoc analysis. First of all, in the speech of many young children, these early auxiliary-like elements form a system: can't is the negative form of hafta, don't the negative form of gonna (see Davis 1983); Hyams' analysis misses this generalization. Second, it is highly implausible that the child should postulate three morphologically distinct variants of the negative particle. Third, Hyams' claim that children have correctly mastered the inflectional endings on hafta and gonna is dubious, to say the least; there is no evidence that at the stage in question children have mastered any inflection endings except the highly productive, phonologically transparent progressive suffix ing, or, more to the point, that they have analyzed semi-auxiliaries into V + inflectional ending.

Third, Hyams' analysis makes the prediction that auxiliaries will emerge at the same time in declarative and interrogative structures. On the basis of Bellugi (1967), Hyams claims that this prediction is borne out. However, even a slightly more extensive review of the literature on the acquisition of auxiliaries (which will be examined in detail in Chapters 3

and 4) shows that this is simply false (see for example Kuczaj and Maratsos 1975, Maratsos and Kuczaj 1983, Fletcher 1979). In Chapter 4 we will investigate the grammar of a child with perfect control of the auxiliary system in declaratives who for many months refused steadfastly to use inverted auxiliaries in yes/no questions.

In short, Hyams' account is flawed metatheoretically, theoretically, and empirically.

Guilfoyle (1984) has proposed an alternative to Hyams' analysis in which the "missing subject" stage is accounted for not by pro-drop, but by a failure of Case-assignment to the subject position. This analysis avoids the problem of an implausible parameter setting, but creates others. In particular, Guilfoyle is forced to acount for the lexical subjects that do occur at this stage as "topics" (following Gruber 1969); but since topics head a Case-marked chain, this is not really a solution at all. Lebeaux (1987) attempts to rescue Guilfoyle's analysis by assuming an entirely different Case-assigning mechanism for children and adults, in which topics can get "structural" Case; however, since there is no independent evidence for this assumption, it appears to be merely an ad-hoc stipulation - and, as mentioned in Chapter 1, a type of stipulation which allows a virtually unconstrained relationship between adult and developing grammars. Moreover, the Guilfoyle/Lebeaux analysis encounters many of the same empirical problems as that of Hyams, most notably the existence of as many "verbless" as "subjectless" sentences at the

developmental stage in question.

Yet another analysis is put forward by Weinberg (1987). Weinberg's account (briefly discussed in Chapter 1) combines elements of performance-based explanations such as Bloom's with Hyams' grammar-based explanation; her claim is that a performance constraint prevents children at the two-word stage from producing SVO sentences (she points out that in general at this stage children can produce SV sentences with intransitive verbs, as shown by Goldin-Meadow 1979); this forces them to set the pro-drop parameter in the "Italian direction" in order to produce grammatical utterances.

Weinberg's analysis is ingenious, but in some ways embodies the worst of both the performance-based and grammar-based explanations. In particular, it appeals to an unanalyzed sentence-span constraint in the same way as Bloom's original complexity hypothesis, and, since Weinberg adopts most of the details of Hyams' pro-drop analysis, it has the same empirical and theoretical weaknesses as the latter.

Let us then turn to a possible alternative treatment of the "missing subject" stage.

First of all, let me point out that all the analyses discussed above, though in some ways diametrically opposed, make a very similar assumption. The assumption in question is what we shall call "the fully articulated" phrase-structure assumption". Thus both Bloom and Hyams attribute fully-articulated phrase-structure configurations to the child, though for rather different reasons: Bloom is assuming that

phrase-structure representations characterize prelinguistic as well as linguistic representations, while Hyams assumes phrase-structure parameters are set very early, and that learning phrase-structure is a prerequisite to learning other syntactic components. Now suppose we drop this assumption. In that case we avoid both the problems attendant on Bloom's analysis (an unspecified theory of performance constraints) and on Hyams' (an implausible parameter-setting). To put it simply, the "missing" subject stage is presyntactic; since at the period in question, children have not yet begun to apply the principles which map strings into trees, we should not <u>expect</u> their language to resemble ours. The "pro-drop parameter" debate is then an acquisitional storm in a teacup.

It will probably objected that this is all very well, but a flawed explanation is better than none at all, and I have simply abandoned the whole explanatory enterprise. However, note that the model I have been developing actually makes some rather strong predictions as to exactly <u>when</u> in the course of development syntactic licensing principles will come into play. In particular, the P-licensing conditions are all dependent on the acquisition of AGR, which in turn depends on the identification of INFL. I thus claim that identifying certain inflectional elements is a precondition - a trigger, if you like - for entering a purely syntactic mode of learning. This locates the "entry into syntax" quite specifically - in English, it will coincide with the learning of the auxiliary system. This hypothesis will be explored in greater detail in

Chapters 3 and 4.

2.96 Conclusion and Summary of Chapter 2

This ends our second chapter. Let us review what has been accomplished.

First of all, we have now introduced the major components of the GB system, some of which will play an important part in Chapters 3 and 4.

Second, we have introduced some major revisions into the theory itself. These include:

(a) Adoption of a "weak" θ-citerion, and consequent abandonment of PRO.

(b) A novel approach to predication, incorporating the concepts of AGR-linking and AGR-domain, which will play an important part in Chapter 4.

(c) A revised theory of the ergative Case-marking parameter.

(d) A radical approach to phrase-structure, incorporating several innovations. These include (i) adoption of the GPSG idea that all non-local trees are constructed out of local trees; (ii) a theory of categorial association based on the interaction of three types of syntactic category: θ-heads, G-heads, and C-heads; (iii) a set of syntactic licensing

conditions, including A-visibility, P-visibility, C-visibility, and G-visibility, which constrain the mapping of categories onto trees; (iv) adoption of a "saturation" approach to the succession condition on X-bar grammars; (v) abandonment of the uniformity condition on phrase-structure.

(e) A new, revized version of the Projection Principle, the Generalized Revized Extended Projection Principle (GREPP).

Third, we have tried to treat GB theory as a particular (abstract) hypothesis about language acquisition. This has involved the introduction of several new ideas, including:

(a) The idea that the notion of epistemological priority, embodying a particular approach to "plausibility in language acquisition" should guide theory construction.

(b) The hypothesis that the child can build up phrase-structure representations by mapping epistemologically primitive components of the system (i.e., Case-theory, θ-theory, and "functor theory") onto trees via the structural relation of government, instantiated in the form of a set of Canonical Government Configurations; and the syntactic corollary of this approach, which is that government is not a unified relation but a set of relations.

Fourth, we have constructed a model of the real-time acquisition of phrase-structure, incorporating an explicit set of percolation principles (ultimately derivable from the licensing conditions) and canonical government configurations.

Fifth, we have compared our approach to two other versions of phrase-structure acquisition, those of Pinker (1984) and O'Grady (1987).

And finally, we have reviewed the "pro-drop parameter" debate in the language acquisition literature, arguing that a simpler and empirically more adequate explanation is available if we drop the "fully articulated phrase-structure hypothesis" adopted by all participants in the debate.

In Chapters 3 and 4 we will narrow the focus of the thesis, concentrating more exclusively on the syntax of the auxiliary system. At the same time, however, we will widen the data-base to include cross-linguistic analysis; our aim is to develop a parametric model which we will then test against data from language acquisition, in order to answer several questions which have emerged from our discussion in Chapter 1.

CHAPTER 3

ON AUXILIARIES

3.0 Introduction

In the first part of this chapter, we will be taking a detailed look at the auxiliary system in English, employing the framework developed in Chapter 2. We will be focussing on the auxiliary system for three reasons. First, ever since the analysis proposed by Chomsky (1957) it has occupied a unique (if paradoxical) role in generative syntax as a testing ground for new theories. Thus, at the same time as many introductory textbooks are using it as an example of a system particularly suited to a transformational treatment, lexical theorists such as Lapointe (1980a.), Gazdar, Pullum & Sag (1982) (henceforth GPS) and Falk (1984) are using the same set of data to argue against such a treatment. Second, the auxiliary system is finite; this means that it is of limited interest to linguists concerned with formal models of language learning, since any finite state automaton could learn it, but all the more interesting to empirically-oriented linguists if, as we shall see, children do not act like finite-state automata in learning it. Third, there is a particularly rich set of acquisition data connected with learning auxiliaries which provides us with a data-base for examining the connection (if there is one) between cross-linguistic markedness and markedness in language

acquisition.

There exists an abundant literature on English auxiliaries; we will not attempt to survey the whole range of analyses put foward at one time or another, though we will attempt to cover the main areas of disagreement. These include the following (as yet not fully resolved) questions:

(i) What is the category status of auxiliary elements ? Is there a separate category AUX (or INFL, in its GB incarnation), as first argued by Chomsky (1957) and subsequently by (amongst others) Akmajian and Wasow (1975), Akmajian, Steele & Wasow (1979) (henceforth ASW), Lightfoot (1979) and Roberts (1985) (from a diachronic standpoint), and Steele et al. (1981) (from a cross-linguistic viewpoint) ? Alternatively, can auxiliary elements be treated as special types of verbs, as originally suggested by Ross (1969) and subsequently by McCawley (1971), Pullum and Wilson (1977), GPS (1982) and Falk (1984), amongst others ?

(ii) What is the nature of the processes known (after Chomsky 1957) as "affix-hopping" and "do-support/do-deletion" ? Are they to be treated as syntactic rules, as argued by most "transformational" theorists (e.g. Emonds 1976, 1985, Jackendoff 1977, Culicover 1976, ASW, and in its GB incarnation, "Rule R", by Chomsky 1981) ? Alternatively, is there a simpler formulation involving subcategorization, as argued by many "lexical" theorists, including McCawley (1971),

Pullum & Wilson (1977), Lapointe (1980a.), GPS, Falk (1984), and Pinker (1984) ?

(iii) Is there a syntactic rule of Subject Auxiliary Inversion (SAI) ? Again, most transformationalist and GB theorists (including all of those mentioned in (ii) above, as well as, within the GB framework, Safir 1982, Koopman 1984, Travis 1984) argue for its existence, and most lexicalist theorists (including GPS, Falk, and Pinker) against. This topic will be dealt with in much greater detail in Chapter 4, and touched on only indirectly in the present chapter.

It will be observed that though these three questions are in principle independent, there are quite clearly two main opposing camps, which we might label "syntactic" and "lexicalist". This is understandable: theories such as Generalized Phrase Structure Grammar and Lexical Functional Grammar which do away with syntactic movement rules are forced a priori to adopt lexicalist positions on questions (ii) and (iii), therefore it is not surprising to find practitioners of such theories in the lexicalist camp. However, GB theorists are perhaps more fortunate in being able to adopt either a syntactic or a lexicalist position (though the former is far more popular); accordingly, we will attempt to treat these questions as far as possible as separate, empirical issues.

The second part of Chapter 3 will be concerned with situating the English system in a cross-linguistic context.

Obviously, both the scope of this thesis and the lack of detailed data comparable to that available in English and related languages will limit the extent of such an investigation; we do, however, at least hope to suggest a framework within which more detailed comparison may be possible; in particular, we will propose a particular set of parameters governing variation in auxiliary systems.

The third part of the chapter will be devoted to an investigation of the acquisition of the auxiliary system in English, with a view to examining the relation betwen the cross-linguistic parametric model developed in the second part and real-time language acqiuisition data.

First however, let us present the basic English facts that any theory must be able to account for. These are succinctly stated in Lapointe (1980a.), on which the following account is partially based.

3.1 Auxiliary Restrictions

3.10 Ordering Restrictions

Auxiliary elements in English occur in an invariant order:

(1) (Modal) - (Auxiliary have) - (Auxiliary be) - (Passive be) - Main V

Each of the parenthesized elements is optional. Thus (1) below

illustrates the only possible ordering of auxiliaries; all twenty three other permutations lead to ungrammaticality:

(2) I could have been being massaged by dancing girls, but instead I ended up in the dungeons.

3.11 Morphological Restrictions

Each auxiliary element triggers a form of agreement on the verb to its right. In any auxiliary sequence, the following restrictions will hold:

(i) The first (leftmost) member of the sequence will appear in ten form (underlined in (3) below):

(3)a. John could win the race.
b. John has won the race.
c. John is winning the race.
d. The race was won by John.

(ii) Following a modal, a verb will appear as a bare stem:

(4)a. John could win the race.
b. John could have won the race.
c. John could be winning the race.
d. The race could be won by John.

(iii) Following auxiliary have, a verb will appear affixed with

-en, the past participle:

(5)a. John has won the race.

b. John has been winning the race

c. The race has been won by John.

(iv) Following auxiliary be, a verb will appear affixed with -ing, the progressive participle:

(6)a. John was winning the race.

b. The race was being won by John.

(v) Following passive be, a verb will appear affixed with -en, the past participle:

(7) The race was won by John.

3.12 Subject Auxiliary Inversion

The first (leftmost) auxiliary element in a string undergoes SAI; this includes main verb be but excludes all other main verbs, including main verb have (except in formal British English). Just in case no auxiliary is available to undergo SAI, auxiliary do is present:

(8)a. Could John win the race ?

b. Has John won the race ?

c. Was John winning the race ?

d. Was the race won by John ?

e. Is John a good runner ?

f. ??Has John good running shoes ?

g. *Won John the race ?

h. Did John win the race ?

3.13 Sentence Negation

The negative element not can appear anywhere to the right of the first auxiliary element and to the left of auxiliary be, though acceptabilty deteriorates the further to the right not is positioned:

(9) I (not*) could (not) have (?not) been (??not) being (*??not) massaged by dancing girls.

When contracted onto a preceding auxiliary, not undergoes SAI:

(10)a. Couldn't John win the race ?

b. Hasn't John won the race ?

c. Isn't John winning the race ?

However, contraction is only possible with the first auxiliary:

(11)a. Could John have not won the race ?

b. *Could John haven't won the race ?

When no auxiliary element is present, auxiliary do is inserted preceding not; as with SAI, main verb be but not main verb have counts as an auxiliary:

(12)a. John did not win the race.

b. *John won not the race.

c. John is not a good runner.

d. *John does not be a good runner.

e. ??John has not good running shoes.

f. John does not have good running shoes.

3.14 Floating Quantifiers

"Floating" quantifiers such as all, both, each appear in the same environments as not, as well as before the first auxiliary:

(13) The men (all) could (all) have (all) been (all) being (*all) massaged by dancing girls.

3.15 Adverbs

S-adverbials (those which can appear in the frame "It is __ the case that" - see Jackendoff 1972) can appear anywhere from the beginning of a sentence to a position preceding auxiliary be:

(14) (Probably) John (probably) could (probably) have (?probably) been (*probably) winning the race.

In contrast, VP (manner) adverbials occur from a position following the first auxiliary to one preceding the main verb:

(15) (*Ruthlessly) John (*ruthlessly) has (ruthlessly) been (ruthlessly) outpacing all his rivals.

3.16 VP-Rules

Several rules in English refer to the "movement" or "deletion" of VP (no theoretical importance is attached to these traditional terms; the important point is that however the constructions in question are to be described, reference must be made to some type of V^n constituent.) In principle, these processes should provide ideal constituency tests for auxiliary elements; however, there is widespread data disagreement in the very extensive literature which has developed around this topic, and judgements appear to be sensitive to various pragmatic as well as semantic factors. We will not attempt to give a thorough survey of the problems involved; for detailed analysis see Ross (1969), McCawley (1975), Akmajian & Wasow (1975), Culicover (1976), Emonds (1976), Sag (1980, 1982), Hankamer & Sag (1976), Hudson (1976), Pullum & Wilson (1977), Iwakura (1977), Williams (1977, 1978), Huddleston (1978), ASW, GPS, Zagona (1982), Falk (1984), Pinker (1984), Haik (1985), Lobeck (1986).

(i) "VP-deletion"

In a configuration such as that in (16), all but the first of a series of auxiliaries may be deleted along with the main verb and its complements (though there is some disagreement as to whether passive be can license an empty VP: see Akmajian & Wasow 1975, Huddleston 1977):

(16) John might have been being beaten, and Bill might (have(been(being(beaten)??))), too.

When no auxiliary is present, auxiliary do appears:

(17) John won a race, and Bill did, too.

(ii) "VP-fronting"

Often analyzed as a sub-case of topicalization, this rule fronts a V^n constituent which cannot include an auxiliary:

(18) I told John he could have been winning, and

a. ...winning he could have been !

b.*...been winning he could have !

c.*...have been winning he could !

Note that main verb be does not behave like an auxiliary in this construction but like main verb have:

(19)a. I told John he was being a fool, and being a fool he

was !

b. I told John he would have a lousy Christmas, and having a lousy Christmas he is !

Again, when no auxiliary is present, auxiliary do appears:

(20) I told John he would win, and win he did !

(iii) Other VP rules

Several other constructions in English involve "empty" VPs. including tag questions, conjunctive so-preposing, and comparatives, which will be discussed in more detail in Chapter 4. It is possible to analyze all three as subcases of VP-deletion, since the same pattern of "auxiliary-stranding" seems to characterize them:

(21)a. Bill could have been being watched by the RCMP, couldn't he (have(been(??being))).

b. Bill could have been being watched by the RCMP, and so could John (have(been(??being))).

c. Bill might have been being observed more intensely than might John (have(been(??being))).

We are thus left with only one "VP-deletion" process which enters into the analysis of all the constructions above. We predict that independent factors will account for the differences between them, a prediction to which we will return below.

Taking a similar approach to exclamative so-preposing, note

that this construction fronts any X^{max} category except VP:

(22)a. So stupid was John that he forgot his own name.

b. Such a fool did John appear that nobody took him seriously.

c. So in love was John that he forgot all his troubles.

d. So quickly did John run that he beat everybody.

e. John is so irritating (that I can't bear to talk to him)!

f. So irritating is John (that I can't bear to talk to him)*!

So, of course, does WH-movement:

(23)a. How many did you see ?

b. How stupid are you ?

c. How quickly do you run ?

d. In which race did you run ?

This suggests that the same fronting process should account for both constructions, and, again, that independent factors account for the differences.

3.17 Imperatives

There appear to be rather strong auxiliary restrictions on imperatives. In particular, modals can never appear; this is not a semantic restriction, as can be seen from (24b.) below:

(24)a. *Can do quadratic equations by Wednesday, or else !

b. Be able to do quadratic equations by Wednesday, or else !

The status of auxiliary have and be is less clear. be seems fine with most imperatives, but have is marginal (see Schmerling 1977, ASW, GPS).

(25)a. Be sitting here when I get back, or else !

b. ?Have finished your homework by the time I get back, or else !

The "why (not)" construction seems to be a subcase of the imperative construction, and auxiliary restrictions appear to be roughly parallel between the two. Thus ASW point out that sentences beginning with why or why not as in (26) below never contain modals. They claim that auxiliary have is also ungrammatical in this construction, a claim disputed by GPS; auxiliary be, on the other hand, is indisputably grammatical:

(26)a. Why (should*) finish your homework over the weekend ?

b. ?Why not have finished your homework by Monday ?

c. Why not be finishing your homework when your father gets in ?

3.18 Complementation

An important source of information about the constituent structure of auxiliaries is to be found in the structure of "defective" complement clauses. These include infinitivals (i),

gerundives (ii), complements to let and make (iii), and complements to perception verbs (iv).

(i) Infinitivals

In English, auxiliaries in infinitivals differ from those in tensed sentences in the following ways:

(i) Nominative case is never assigned:

(27)a. I want to leave.
b. I want (for) him to leave.
c. *I want I/he to leave.

(ii) Modals are impossible:

(28)a. I hoped I could go.
b. *I hoped to could go.

(iii) Not appears before the infinitival marker to:

(29)a. It was believed that he (*not) had (not) left.
b. He was believed (not) to (??not) have (not) left.

(iv) Like modals, infinitival to cannot be deleted by VP-deletion:

(30) Mary wants to leave early, and John wants (to)* as well.

(v) to contracts onto a preceding verb just in case no lexical NP or trace in argument position intervenes between the two. The "wanna contraction facts have been hotly debated: see Jaeggli (1980), Postal & Pullum (1982), Pesetsky (1982), Bouchard (1984), amongst others. The basic contrast is as in (31) below:

(31)a. Who do you wanna see t ?

b. *Who do you wanna t see Bill ?

(ii) Gerunds

Gerundive ("poss-ing") complements in some ways closely resemble infinitivals. They assign no case to subject position:

(32)a. I tried going.

b. *I tried him going

Exceptional Case-marking verbs which take poss-ing complements can thus assign accusative Case to an embedded subject, as in (33b.); (33c.) illustrates a nominal gerundive complement, which being an ordinary NP, assigns genitive Case to its subject.

(33)a. I hated going.

b. I hated him going.

c. I hated his going.

Not attaches to the front of the gerund:

(34)a. I tried not going.

b. I hated him not going.

c. *I hated his not going.

Modals are impossible:

(35)a. I hated not being able to finish my thesis.

b. *I hated not can(ing) finish my thesis.

A difference between infinitivals and gerunds lies in the latters' inability to take progressive be:

(36)a. I tried not to be looking out of the window when the teacher arrived.

b. *I tried not being looking out of the window when the teacher arrived.

This contrast has usually been accounted for by some version of the "Double-ing" constraint (Ross 1972, Reuland 1983). Auxiliary have occurs freely in gerunds, as in infinitives:

(37)a. I tried to have finished my homework when the teacher arrived.

b. I tried having finished my homework when the teacher

arrived.

(iii) Complements to let and make

These verbs take "bare" infinitives, though note that when passivized, an infinitival to surfaces with make, whereas the passive equivalent of let is ungrammatical:

(38)a. I made him leave.
b. I let him leave.
c. He was made (to)* leave.
d. *He was let (to) leave.

Modals are completely ungrammatical:

(39)a. *I made him will leave.
b. *I let him could leave.

Results are mixed with auxiliary have and be. Be appears to be reasonably acceptable and have reasonably unacceptable, but judgements are not sharp. ASW claim the following pattern of grammaticality:

(40)a. *I made him have finished his homework by the time I got home.
b. I forced him to have finished his homework by the time I got home.
c. *I let him have finished by the time his girlfriend arrived.

d. I allowed him to have finished by the time his girlfriend arrived.

e. I made him be doing his homework when his favourite TV programme was on.

f. I let him be watching TV when his mother arrived.

Not attaches (somewhat marginally) to the bare verb complement:

(41)a. I made him not go to to the movies.

b. ?I let him not go to the movies.

The complement is a possible target for VP-deletion, just in case an auxiliary is available to "support" the stranded tense; Note that auxiliary do, like modals, is impossible:

(42)a. I made John be late, but I didn't make Bill (be).

b. I made John stay awake, but I didn't make Bill (*do).

c. I let John be arrested, but I didn't let Bill (be).

d. I let make up his own mind, but I didn't let Bill (*do).

(iv) Verbs of perception and verbs of temporal aspect.

Like let and make, these verbs take bare infinitival complements. Passivized perception verbs, like make, surface with infinitival to:

(43)a. I heard him leave.

b. He was heard (to)* leave.

c. I heard him leaving.

d. He was heard leaving.

Verbs of temporal aspect (start, stop, keep, continue, etc.), which take gerundive complements seem to pattern with perception verbs in allowing neither auxiliary have nor auxiliary be:

(44)a. *I heard him having left.

b. *I heard him being making a noise.

c. I began making a noise.

d. *I began having made a noise.

e. *I began being making a noise.

Emonds (1976) claims that the ungrammaticality of examples such as (44d.) and (e.) can be attributed to semantic anomaly, since aspectual auxiliaries are semantically incompatible with "aspectual" verbs. Though this analysis is intuitively appealing, there seems to be an appreciable improvement in acceptabilty if the same verbs take infinitival complements:

(45)a. ??I began to have left by the time he arrived on Tuesdays.

b. ?I began to be sitting at the table whenever he came in.

Though there is no doubt a contributing semantic factor, there remains a residue of syntactic behaviour to be explained.

Negation of complements of these verbs is possible only with "constituent negation" (see 3.43 below):

(46)a. ?I heard him not be quiet.

b. ?He began not going out of the house.

VP-deletion is generally impossible (pace the claims of ASW who maintain that it is possible only if main verb be is available to support stranded tense):

(47)a. *I saw him be humiliated by his girlfriend, and John saw him (be) too.

b. *I saw him be a fool, and John saw him (be), too.

3.19 Other Constituency Tests

"Right node raising" has often been used as a constituency test (though see Abbot (1976) for an unexplained counterexample). Applied to English auxiliaries, we get the following pattern of grammaticality (the "raised" constituent is underlined):

(48)a. John might have been being, and I should have been being, watched by the police.

b. John might have been, and I should have been, being watched by the police.

c. ?John might have, and I should have, been being watched by the police.

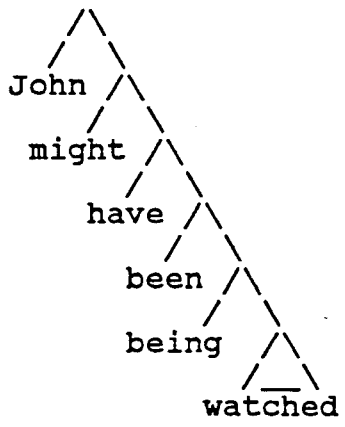
d. ?John might, and I should, have been being watched by the police.

e. John, and Bill, might have been being watched by the

police.

This indicates the following constituent structure:

(49)



This is supported by another constituent test, that provided by parentheticals. Emonds (1976) makes the interesting claim that any material to the right of a parenthetical will form a constituent. Assuming this to be a valid generalization, let us apply it to the structure of the English auxiliary system. Again, the relevant "constituency candidate" is underlined.

(50)a. John, I think, might have been being watched by the police.

b. John might, I think, have been being watched by the police.

c. John might have, I think, been being watched by the police.

d. John might have been, I think, being watched by the police.

e. ?John might have been being, I think, watched by the police.

Though there is some variation in acceptability in these

examples, probably due to the "focussing" effect of rightward movement (see Rochemont 1986), they can all be rendered grammatical by provision of a suitable context. We conclude that the constituent structure proposed above is the correct one. Notice, however, that we have not yet addressed the category status of the various auxiliary elements. This has proved to be a far more contentious issue, to which we now turn.

3.2 On the Category Status of Auxiliary Elements

3.20 Modals

Modals can be distinguished by the following properties:

(i) Modals never cooccur, and always occur first in a string of auxiliaries (3.10).

(ii) Modals can never be morphologically suffixed (3.11).

(iii) Modals always invert in SAI constructions (3.12).

(iv) Modals always precede not in sentential negation (3.13).

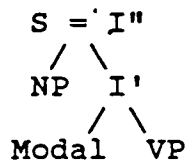
(v) Modals never delete in VP-deletion environments, nor prepose in VP-fronting environments (3.16).

(vi) Modals never appear in imperatives, or in the why not construction (3.17).

(vii) Modals never appear in any "defective complement" environment (3.18).

These facts would appear to constitute strong evidence for treating modals as a separate category from main verbs, which have none of the properties outlined above. This has led to the "classical" view of the English auxiliary system, first proposed by Chomsky (1957), in which modals are not verbs, and are not generated under the VP node, but under a separate AUX (INFL) node. This leads us to postulate structure (51) below:

(51)



The alternative view, advocated by many of the "lexicalist" camp (see references in 3.0 above), is that modals are special cases of the category [V]. The arguments for such a view are based on the following considerations:

(i) There is no sharply defined class of modal elements. While "core" modals behave as predicted by the criteria

above, many elements share only some of these properties. These include the semi-auxiliaries need, dare, ought, and used. In addition, there is dialectal variation in modal behaviour: for example, in some Texas dialects "double modal" constructions are common (see Fabb 1983).

(ii) Modals share many of the properties of verbs: they absorb tense and undergo "affix-hopping"; they may be gapped; they behave like verbs for the purposes of the double-ing filter; and they assign θ-roles in exactly the same way as ordinary verbs.

(iii) Alternative analyses are already available which predict many of the properties associated with modals and are independently necessary for the analysis of uncontroversial main-verb structures, thus rendering a separate "modal" category redundant. For example, as we have seen, there are non-modal verbs such as make and let which take bare VP complement structures; let, in addition, fails to passivize and never appears with auxiliary to, both properties usually associated with modals.

We will argue that while some of these arguments may be correct, none of them necessitates collapsing the categories [I] and [V], and such a move would lose the generalization that modals do share a cluster of properties not associated with

main verbs. The lack of clear-cut categorial distinctions between [I] and [V] elements will follow from two separate assumptions. Firstly, as argued in Chapter 2, [I] and [V] are intrinsically associated via P-visibility and G-visibility; the nature of this association will be discussed in detail in 3.4 below. Secondly, we will adopt from Emonds (1985) the notion of Unique Syntactic Behaviour, which predicts that closed-class items will differ syntactically not only from open-class categories but from each other; this will be discussed in greater detail in 3.3.

Provisionally, then, let us conclude that modals belong to a category [I], distinct from the category [V].

3.21 Auxiliary Do

Turning to auxiliary do, we can make the following generalization: do appears in AUX when

(i) there is no other auxiliary present, and

(ii) tense is "separated" from a main verb under conditions yet to be specified, or

(iii) the sentence containing do has an "emphatic" reading.

The exact conditions of (ii) have been one of the focuses of the auxiliary debate. Chomsky's (1957) account introduced the

transformation of "Affix Hopping", which (informally) suffixes an abstract tense morpheme to an adjacent [+V] element, where [+V] includes both modals and ordinary verbs. When the two are non-adjacent, the dummy verb do is inserted in order to ensure that the tense suffix is attached to some lexically overt [+V] category. This will account for the presence of do in SAI constructions (3.12), negated sentences (3.13), VP-deletion environments (3.16), and for its absence in the why (not) construction (3.17) and in defective complement environments (3.18).

An alternative to do-support, do-deletion, has been proposed by Culicover (1976), Emonds (1976), and adopted by ASW. This is essentially the converse of do-support, and involves the presence of do at D-structure and its deletion in the syntax just in case it precedes a [+V] element. The purported advantage of such an account is that it obviates the need for a disjunction in the statement of SAI, which must otherwise refer to a sequence of [Tense] + (Modal or have or be).

Note that neither do-support nor do-deletion will account for the presence of do in imperatives such as

(52)a. Do be a good boy !

b. Do have another drink !

One way of generating such structures would be to derive them from emphatic do; note, however, that (53a.) - but not (53b.) - is ungrammatical:

(53)a. *He did be a good boy !

b. He did have another drink !

(53) shows that it is a peculiar property of be rather than of auxiliary do which is responsible for the contrast between (52a.) and (53a.); we will return to a more detailed discussion of be in the next section. Note that here the parallel distribution of modals and auxiliary do breaks down; modals are ungrammatical in imperative environments (see 3.17 above).

Do is fully inflected morphologically, in contrast to modals, which have defective morphology. This may be linked to the fact that do also functions as an ordinary main verb; note that auxiliary has, which also functions as a main verb, also has fully inflected morphology.

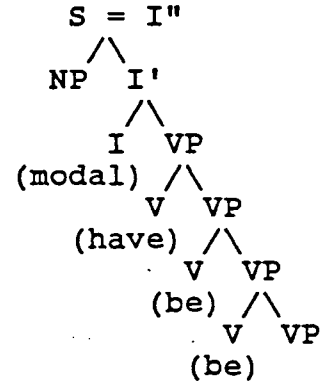
Thus, out of the seven criterial modal properties listed above, do shares five. In view of this similarity, we could treat do simply as a semantically "neutral" modal, a member of the category [I]. We will argue that such a view is basically correct, and that the differences between do and modals in fact derive from separate levels of lexical insertion, rather than categorial divergence.

3.22 Auxiliary Have and Be

Turning to auxiliary have and be, we see that these aspectual elements behave like modals and do with respect to (2.312,

2.313, 2.316) as long as they are the first element in a string of auxiliaries. In order to capture the dual nature of non-modal auxiliaries, several authors (Culicover 1976, Emonds 1976, Jackendoff 1977) have proposed that they be generated as verbs under VP but undergo a rule of have/be raising into AUX, just so long as no lexical auxiliary already occupies the AUX slot. Thus prior to raising, aspectual auxiliaries would occur in a structure such as that below:

(54)

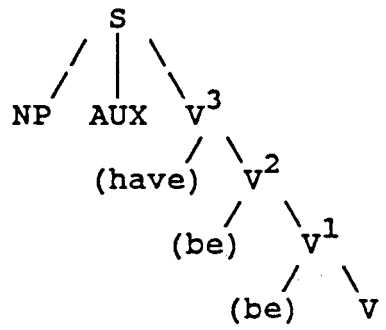


We will adopt the idea that this rule is the analogue of a more widespread rule of verb-raising which at one point also applied to all English verbs (Roberts 1985). In order to ensure that only aspectual have and all forms of be may undergo verb-raising, appeal is generally made to a [+AUX] feature; whether such a stipulative feature may be dispensed with in favour of a more explanatory generalization will be discussed below.

ASW, in their exploration of the subtler aspects of the auxiliary system, adduce a number of grammatical phenomena purporting to justify a much more elaborate, "layered" structure for VPs containing have and be. They distinguish

three distinct VP levels, as in the structure below:

(55)



V^3 contains aspectual have; V^2 progressive be; and V^3 passive be. ASW claim that the grammar of English contains enough evidence for a child to be able to distinguish all three levels.

Evidence for V^3 is to be found in any construction which does not permit modals (or auxiliary do) but does permit aspectual have and any element in the verbal projections below it. Infinitival and gerundive complements (3.18) provide such evidence.

Evidence for V^2 is less convincing; ASW argue that imperatives (3.17), the why (not) construction (3.17), and complements to let and make (3.18) permit progressive be but not aspectual have. GPS, on the other hand, argue that the constraints on aspectual have in these constructions are semantic, all involving difficulty in construing perfective aspect with a "deontic" sentence modality involving acts of permission or exhortation. They further point out that deontic may - clearly a modal - suffers from the same difficulties of interpretation with perfective have:

(56)a. John may drink up to three bottles of Scotch.

b. John may have drunk up to three bottles of Scotch.

c. John is allowed to drink three bottles of Scotch.

d. John was allowed to have drunk three bottles of Scotch.

e. John may be drinking three bottles of Scotch.

(56a.) is ambiguous; the modal may either have a "root" (deontic) interpretation, in which case may can be paraphrased by is allowed, as in (56c.); or it may have an epistemic interpretation, in which case it may be paraphrased by "...it is possible that". (For an analysis which exploits the parallelism between these two interpretations and the interpretations assigned to S- versus VP- adverbials (Jackendoff 1972) see Zubizaretta 1982). (56b.), in contrast, permits only the epistemic reading, indicating the incompatibility of perfective aspect with deontic interpretation. (56d.) however, casts doubt on a purely semantic account of (56b.), since it is grammatical and a paraphrase of (56b.); and (56e.) shows that the deontic reading for a modal is blocked by any aspectual marker, including progressive be, which is permitted to appear in the environments specified above where have may not appear. We thus have to conclude that a purely semantic account of "V^2" is inadequate. Yet such an account appears to be on the right track, since given suitable contextual support, semi-grammatical sentences can be constructed with auxiliary have

and two of the three "V2" constructions - namely, imperatives, as in (57a.), and why(not) sentences, as in (57b.):

(57)a. Have finished your homework by the time I get back or else !

b. Why have finished your homework by the time he gets back ? He'll only give you more.

Complements to make and let, however, seem to be simply ungrammatical with aspectual have; near paraphrases with force and allow are not nearly as deviant:

(58)a. *He made me have finished my homework by the time he got back.

b. He forced me to have finished my homework by the time he got back.

c. *He let me have finished my homework before assigning a grade.

d. He allowed me to have finished my homework before assigning a grade.

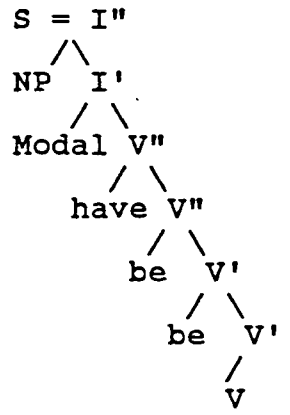
Though such evidence seems to point to the existence of a syntactic rather than semantic difference between these two sets of verbs, as Lapointe (1980b.) points out, the difference seems to hold up cross-linguistically; this is unexpected if we are dealing with idiosyncratic aspects of subcategorization, but not if the two sets of verbs are semantically distinguished. We conclude that though no adequate semantic account currently exists for the selectional restrictions which ASW take as evidence for the V^2 level, ultimately such an account holds out more explanatory promise than a purely

syntactic treatment.

Turning to V^1, ASW cite VP-preposing (3.16), complements to perception verbs (3.18), and complements to verbs of temporal aspect (3.18) as evidence for a VP level including passive <u>be</u> but excluding other auxiliary elements. We find such evidence for a "minimal" VP uncontroversially convincing.

We thus conclude that two out of the three levels postulated by ASW seem well-supported by the evidence, but that the existence of the V^2 level seems more dubious. Casting these conclusions into a conventional X-bar framework, we come up with the following structure:

(59)



In the next section, (59) will be amended to fit our rather different conception of phrase-structure.

Returning to the categorial status of auxiliary <u>have</u> and <u>be</u>, we conclude - along with virtually all work since the "<u>have/be</u>-raising" analysis was first proposed - that these elements are verbs. This is supported by the fact that they show none of the seven criterial properties we have associated with modals.

Nevertheless, they are not ordinary verbs, from which, as we have seen, they differ in the following ways:

(i) They may under certain circumstances invert.

(ii) They may under circumstances occur in pre-NEG position.

(iii) They may under certain circumstances control a VP-deletion site.

(iv) There are restrictions on their occurrence in imperatives and the why (not) construction; these may or may not be due to semantic factors.

The first three of these differences are accounted for by have-be raising. But a larger question now arises: what accounts for have-be raising ? It is to this and related questions that we now turn.

3.3 On the Association Between [I] and [V]

3.30 Visibility

In this section we will present an analysis of the data presented in the first part of the chapter, arguing that many of the categorial quandaries raised there can be accounted for in the phrase-structure framework developed in Chapter 2.

In Chapter 2 we introduced a conception of phrase-structure which systematically distinguished three types of categories: θ-heads, G-heads, and C-heads.

Three important visibility conditions, A-visibility, P-visibility and G-visibility, constrain the relationship between G-heads and θ-heads. These are repeated below in (60):

(60)a. An A-chain must be associated with Case.

b.I. Every predicate must be contained within an AGR-domain

II. Every AGR-domain must contain a Predication Subject.

III. Every predication subject must be AGR-linked.

c. G-heads must be θ-associated.

d. C-heads must be G-associated.

The conditions in (60) impose an interlocking set of constraints on θ-related elements. The condition of direct interest in this chapter is G-visibility; in the next chapter we will mainly be concerned with P-visibility. Accordingly, let us examine G-visibility more closely.

3.31 G-visibility, Full Interpretation, and [I,V] Association

First of all, we will narrow down our field of inquiry. G-visibility as presented in Chapter 2 is a very general condition on categorial association, but it can be realized in different ways even in the same language. In English, there are

two categorial relations mediated by G-visibility, [D]/[N] association and [I]/[V] association; though we have claimed these associations are functionally parallel, and thus, by hypothesis, configurationally parallel, association is realized differently in each case: [I] elements are bound morphemes, while [D] elements are probably best analyzed as phrasal clitics. Cross-linguistically, we know that the reverse situation can hold; [D] elements may be bound (as in Hungarian) and [I] elements may be free (as in Warlpiri). In this chapter (and indeed, in the rest of the thesis) we will be concentrating on [I]/[V] association, and more particularly, on the type of [I]/[V] association exemplified in English and related languages; this is unfortunate, but inevitable, given the limited nature of this study. In order to broaden the scope of inquiry, I can only make the usual time-honoured appeal to "future research".

Let us then narrow G-visibility down to "I-visibility", in which case we will get the condition in (61):

(61) I-visibility: an I-head must be V-associated.

Our next question concerns the locus of condition (61). Is it an Lf condition, or a PF condition, or a purely syntactic condition ?

Chomsky (1986a.) introduces a very general condition on syntactic elements of all kinds, the Principle of Full Interpretation (PFI), which is repeated in (62) below:

(62) Every element of PF and LF must receive an appropriate interpretation.

While (62) might appear like a truism, it actually imposes a rather radical perspective on the notion of "licensing condition" - the idea that all syntactic principles are derived from conditions on either PF or LF. If we accept this premise, then no purely syntactic (S-structure) conditions will exist. Now, at present, the PFI is more of a research-generating strategy than an empirically supported generalization (in other words, a "leading idea"); however, since it in principle restricts the range of possible licensing conditions, we will adopt it as a working hypothesis. (Note that adopting the FI does <u>not</u> commit us to a view that syntax is trivially related to interpretation; PF and LF are themselves autonomous <u>syntactic</u> levels, and as such, principles applying to them are syntactic in nature).

Now, if we adopt the PFI, a condition such as I-visibility can only have two sources: it is either a condition on PF, or on LF. The latter appears unlikely (though we cannot rule it out on principle) since there seems no semantically-based reason why [I]-elements should have to be associated with [V]-elements; the kinds of semantic feature encoded by [I] (tense, aspect and modality features) are usually treated as propositional operators, with no intrinsic relation to thematic properties of [V] (though see Pollock 1987, who argues that V-

raising is motivated by an operator-variable relation involving tense).

On the other hand, there is considerably more intuitive appeal (and empirical content, as we shall see) to the idea that [I] elements must be linked to [V] elements at PF; this is one way of capturing the common claim that inflectional elements are "bound morphemes". However, note that not all [I] elements are in fact bound to verbs; the [-TENSE] inflectional element <u>to</u> is an independent morpheme with a different ontogeny (it is clearly related to prepositional <u>to</u>) and a partially different syntactic distribution; for example, it intervenes beteen a verb and <u>not</u>. This suggests that if we are to treat [I]-visibility as a PF condition, the "bound element" referred to should not be [I], the G-head, but AGR, which forms a feature complex in tensed [I]. Intuitively, this makes sense: AGR is inherently a syntactic "connector" between subject and predicate. We therefore refine [I]-visibility further to the "AGR-visibility" condition in (63):

(63) AGR must be V-associated.

(63) is a particular instantiation of (61), which is in turn a particular instantiation of the general G-visibility condition in (60c.).

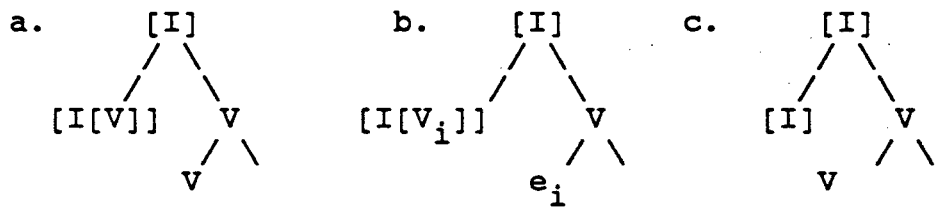
Now, if AGR-visibility is a PF condition, it will have three potential levels of application: at PF, at S-structure, or in the lexicon. It is to the question of level of application that

we now turn.

3.32 Levels of [I]-[V] association.

We have suggested that association between the G-head [I] (or more precisely, AGR-features in [I]) and the θ-head [V] could take place at any of three levels: in the lexicon, in the syntax, or at PF. S-structure configurations corresponding to these levels of association are given below:

(64)



(64a.) is the case of lexical association, where an inflectional element is associated with a verb in the lexicon; (64b.) the case of syntactic association, where a verb attaches to an inflectional element in the syntax, leaving a trace in VP; and (64c.) the case of PF association, where at S-structure [I] and [V] have not yet been associated.

Note that we have not yet discussed the nature of the "derived" categories in (64a.) and (64b.); in our model this will constitute an important source of parametric variation, to be discussed in more detail in section 3.7. There we will suggest that languages may vary as to the "headedness" of the [I,V] category; whereas English is "I-strong", German, for example, is "V-strong" and generates [I,V] constituents in VP.

We shall show that a range of parametrically related properties falls out from this distinction.

However, for the moment, let us stick with English, and the question of levels of [I,V] association. At this point, I would like to claim that we can resolve the controversy concerning the categorial status of auxiliary elements in English by exploiting the three different possible levels of association provided by AGR-visibility. In the next three subsections, I will show that English uses all three levels: PF-association accounts for tensed verbs in declarative sentences; syntactic association gives rise to "have/be raising", the restricted English version of V-raising; and lexical association provides us with modals. Let us examine each of these cases more closely, beginning with PF association.

3.33 PF Association

PF association is forced by the AGR-visibility condition applying at PF, and appears to be the unmarked case for [I]-[V] association. (As we shall see in 3.6, this also applies to languages with "unrestricted" V-raising). This is perhaps to be expected, if we adopt the "last resort" theory of movement rules in which syntactic movement only takes place when some interaction of grammatical principles forces it to occur. In English, PF association fails to occur in tensed clauses when

(i) SAI is forced, or

(ii) Not intervenes between [I] and [V], or

(iii) A modal occupies the [I] node.

(i) is easily explained if, as we assume, SAI is a syntactic movement rule which takes place in the mapping between NP-structure and S-structure. This in turn will follow from the fact that SAI is forced by P-visibility, an LF condition, as we shall argue in Chapter 4. In that case, [I] must already be attached to some [V] (either lexically or syntactically) in order to feed SAI; PF-association does not meet this condition, therefore it cannot take place in such circumstances.

Turning to (ii), appeal is usually made to some type of syntactic adjacency in accounting for the failure of PF association to occur across an intervening not. We will assume that not is an "opaque element" at PF, thus forcing syntactic or lexical association; note that this is an idiosyncratic property of not, not shared with other negative elements, as shown by (65):

(65)a. I don't like to waste my weekends at the keyboard.

b. I never like to waste my weekends at the keyboard.

The third case where PF-association fails to occur falls out straightforwardly from our treatment of modals, which we analyze as cases of lexical [I]-[V] association. Obviously, if [I] is already morphologically bound, PF-association will not take place.

3.34 Syntactic Association

Let us then turn to syntactic association. As we have already mentioned, we treat the rule of have-be raising as a restricted case of the more widespread process of V-raising. It was argued in the last section that this rule was an "NP"-type rule, which left a trace at NP-structure, the level of application of phrase-structure principles.

Have-be raising is "forced" when PF-association is blocked (see 3.33 above). However, it takes precedence over do-insertion (see 3.36 below), suggesting that of the two processes, it is the less marked. We will argue in 3.36 that do is maximally underspecified, so that have-be raising may occur without actually needing to delete any syntactic features. In this sense, it is a "structure-preserving" process.

Perhaps the most salient property of have-be raising is its restriction to only two auxiliary verbs. This is particularly striking in comparison to the equivalent process (V-raising) in French, which applies to any tensed verb. In order to make the required distinction between auxiliaries and other verbs, most analyses of English have resorted to the equivalent of a rule feature [±AUX]; this is hardly illuminating, if mechanically adequate. The real question is one of learnability - that is, how does the child restrict the process to the auxiliary verbs in English, but successfully generalize it to all verbs in the other languages in question ? We will return to this question

in 3.6, when we discuss the hypothesis of Unique Syntactic Behaviour proposed by Emonds (1985).

3.35 Lexical Association

Lexical association, as in (64b.) above, characterizes the class of modals in English. Modals are inserted at NP-structure, but leave no trace, as opposed to aspectual auxiliaries. We will assume (following Culicover 1976) that they are marked [±PAST] and thus absorb tense-features from [I].

This treatment of modals has advantages over both an account which treats them as a separate category and one which treats them as ordinary verbs. Each of these accounts has drawbacks: the latter cannot explain the cluster of properties which differentiate modals from main verbs (see 3.21) while the former suffers from diachronic implausibility, in that it forces historical linguists such as Lightfoot (1979) to propose that English "acquired" a new category (call it "M"), when it lost the productive rule of V-raising.

Our proposal suffers from neither of these drawbacks; modals are differentiated from ordinary verbs by their level of association with [I], but are still verbs for all that: they assign θ-roles,

Roberts (1985) puts foward a proposal which resembles ours in that modals are treated as derived [I]-[V] categories. His proposal differs from ours in several important ways, however. In particular, he makes the following claims:

(66)a. Modals assign no (primary) θ-role.

b. V assigns θ-roles iff V is governed.

Roberts claims that modals appear in ungoverned positions; this follows from the interaction (66a.) and (66b.). Since the only ungoverned position available is INFL, the appearance of modals in INFL is predicted.

However, it turns out that this argument is circular, since the failure of modals to assign θ-roles (66a.) cannot be derived from epistemologically primitive principles. (66a.) cannot be a semantic property of modals (if it were, then the synonymy of <u>can</u> and <u>be able</u> would be inexplicable, as would the ability of modal verbs to assign θ-roles in other languages) but is itself derived from the exceptional <u>syntactic</u> behaviour of modals. In other words, it is assumed that modals assign no θ-roles because they appear in ungoverned positions, and that they appear in ungoverned positions because they assign no θ-roles.

This is precisely the type of circularity which is problematic from the viewpoint of the language-learner, and thus also from the viewpoint of a grammatical theory which takes language acquisition seriously. Our own proposal does not suffer from this circularity; in our model, modals do not differ in θ-assigning properties from other verbs, but only in the level at which they are attached to INFL. We will return to the question of learnability and learning in more detail in

3.8, where we will make explicit proposals as to how the auxiliary system might be learnt, and in 3.9, where we will examine some data from "real-life" acquisition of the auxiliary system.

3.36 Do and Levels of Lexical Insertion

In 3.22 we claimed that auxiliary do resembles the modals in its syntactic behaviour. We might then simply assimilate the distribution of do to that of the modals as a case of lexical [I]-[V] association.

This explanation, however, would not account for the ungrammaticality of the following examples:

(67)a. *I did not have gone.

b. *I do not be going.

What (67) appears to show is that do is only "inserted" after have-be raising has had a chance to apply, or alternatively, that have-be raising "deletes" do; these are of course, the considerations which led to do-support and do-deletion, respectively.

The problem in fact boils down to the question of levels of lexical insertion. In a "monostratal" model where lexical insertion takes place at one level only, do-deletion is the only option. The rule of do-deletion has some suspicious characteristics, however: it requires tense to form a separate

constituent from do at the moment of its application, which in turn inevitably leads to an "affix-hopping" analysis in order to attach tense (now a syntactic constituent, not a feature) to a verb. But if affix-hopping follows lexical insertion, then some very powerful morpho-phonological rules are going to be needed to change (/do/ + [PAST]), for example, into [did].

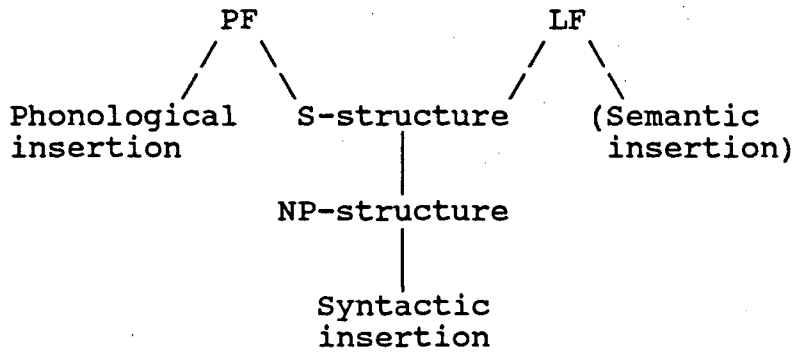
Alternatively, we could formulate a model in which lexical insertion took place after do-deletion. In that case, "deletion" would merely involve replacing the distinctive syntactic features of do with those of either have or be. This would clearly be preferable: syntactically have, be, and do are closely related. Moreover, we could now specify the link between the ability of do to delete and its syntactic "underspecification"; if do were simply not specified for any subcategorization features, then have or be would be able to occupy its "slot" prior to lexical insertion. Obviously, we would have to claim that the phonological insertion of do is a "last resort"; this does not seem an unreasonable hypothesis.

Now we come to an apparent paradox. In Chapter 2 we proposed that subcategorization relations were determined at NP-structure, and that therefore lexical insertion had to take place at that point. Now, however, we are arguing that lexical insertion takes place after NP-structure, in order to accomodate the insertion of do.

The resolution of this paradox lies in abandoning the assumption that lexical insertion takes place in a unitary fashion. While syntactic reasons led us to postulate NP-

structure as the level of insertion for syntactic subcategorization features, morphophonological reasons led us to postulate a post-syntactic level for do-insertion. Suppose we were to separate these two types of "lexical insertion", allowing syntactic features to be inserted at NP-structure but morphophonological features at PF. Under this conception, we would have "syntactic lexical insertion" and "morphophonological lexical insertion", both acting as filters on permissible lexical it but at different levels, as illustrated in (68) below:

(68)



Note that we have included a level of "semantic insertion"; if there is a level of "semantic representation" distinct from LF, then presumably lexical properties of e.g. "idiosyncratic quantifiers" will be encoded there. Note also that we have abandoned D-structure; the claim embodied in this move is that there is no need for a level of representation separate from the subcategorization information contained in the syntactic lexicon.

3.4 Irregularity, Lexical Insertion, and the Designation

Convention

It has often been noted in the "anti-movement" literature (Pullum & Wilson 1977, Huddleston 1978, Schachter 1980, Lapointe 1980a., GPS, Falk 1984, Pinker 1984, to name a few) that various lexical idiosyncracies distinguish auxiliary elements in English; these include the following types of irregular behaviour:

(a) Not all auxiliary elements can occur in all the positions predicted by a movement analysis of the auxiliary system. Thus modal dare and need are unable to occur in ordinary declarative sentences:

(69)a. Dare I eat a peach ?
b. I dare(n't)* go out in the woods today.
c. Need I say more ?
d. You need (not)* apply for this position.

Another notorious example of idiosyncratic syntactic distribution is provided by the first person of negated contracted be, whose distribution is limited to inverted contexts:

(70)a. I am/*aren't going to the movies.
b. Am/aren't I going to the movies ?

Note that in some dialects of English this skewed distribution is "levelled" by use of the contracted negative auxiliary ain't:

(71)a. I ain't a good syntactician.
b. Ain't I a good syntactician ?

(b) The same I-element can have different meanings in different positions: for example, may has two interpretations (deontic or "root" and epistemic or "non-root"). However, only the former interpretation is available in inverted contexts:

(72)a. It may rain tomorrow.
b. *May it rain tomorrow ?

Compare this behaviour with that of might:

(73)a. It might rain tomorrow.
b. Might it rain tomorrow ?

Compare also the similar contrast holding between shall (no epistemic reading in inverted contexts) with will (epistemic reading in any position).

(c) The morphophonological relationship between positive and negative auxiliaries is often suppletive. Thus the alternations will/won't, can/can't, do/don't are morphophonologically

opaque; Zwicky and Pullum (1983) use this as one criterion to distinguish between "inflection" and "cliticization", arguing that at the point of lexical insertion positive and negative auxiliaries are different lexical items.

This is by no means an exhaustive list: see the literature cited above for other examples.

What these irregularities seem to indicate (and as argued in the anti-movement literature) is that occurrences of auxiliary elements in different syntactic positions are ocurrences of separate lexical items. The question is, of course, can we account for such irregularity without resorting to a separate categorial specification distinguishing each lexical item which shows syntactically distinct behaviour ?

I will argue here that two features of the model we have adopted here will allow us to give a positive answer to this question. The first is our treatment of "lexical insertion", which separates syntactic, semantic and phonological lexicons; the second, our adoption from Emonds (1985) of the Designation Convention, which predicts the exceptional syntactic behaviour of closed-class grammatical items. Let us deal with these one at a time.

We have assumed separate syntactic, semantic and phonological lexicons (see (68) above), the latter governing the insertion of semantic and phonological features at LF and PF respectively and effectively acting as lexical filters on well-formed syntactic structures. Thus, if we assume the

inflectional paradigms of morphologically defective elements are listed in the phonological lexicon, then we can filter out syntactically impeccable but ungrammatical sentences such as (69b.) and (70a.) since no appropriate (phonological) lexical item exists to be inserted for the particular syntactic feature-matrix which exists at PF. Similarly, we assume idiosyncratic semantic behaviour to be encoded in the semantic lexicon at LF; if no epistemic reading for <u>may</u> exists in inverted position, (as in (72b.)) then the semantic lexicon will filter it out. (In this sense, different readings will correspond to different "semantic lexical items").

Of course, this implies that both semantic and phonological lexicons can distinguish between negated and non-negated and inverted and non-inverted auxiliaries. But this is no problem: at S-structure, these distinctions exist. We simply need to define phonological and semantic "lexical insertion rules" as context-sensitive, a move generally assumed for lexical insertion rules since Chomsky (1965). For example, first-person <u>aren't</u> will be inserted by a PF lexical rule such as (74) (many irrelevant details left out):

(74) [<u>aren't</u>] / $[_I[_V___]]$ IP

The only configuration corresponding to (74) will be an inverted context.

Thus we see that the "split-lexicon" approach enables us to deal with morphophonological and semantic irregularity without

needing to treat such irregularities as syntactic. However, a deeper problem still underlies this approach. This is the problem of why it should be that certain lexical items typically show "irregular" behaviour, while others do not. To put it another way, if we construct a grammar so that phonological and semantic irregularities are permitted, why should any regularities exist ? As we have argued in Chapter 1, this is in fact a problem which besets most lexical theories, which sacrifice generalizations for exceptions and then have no way of accounting for the pervasive regularities of natural language.

We will adopt a solution to this problem offered by Emonds (1985), who gives a cogent analysis of the empirical reductio-ad-absurdum implicit in attempts to find a unique syntactic criterion for distinguishing one closed-class category from another. Emonds argues that unique syntactic behaviour is actually a defining criterion of closed-class grammatical items (G-heads and C-heads, in our terms), and therefore to be expected. He proposes the following principle to capture this generalization:

(75) Designation Convention: a rule of syntax (whether of insertion, movement, deletion, or filtering) stated in terms of a subcategory of a head of a phrase $[X^0,F]$, where F is a syntactic feature, cannot affect any lexical items under X^0 associated with a purely semantic feature.

The Designation Convention predicts that closed-class items may show irregular behaviour (stated in terms of syntactic features) but that open-class items may not, since open-class items are differentiated from each other only by purely semantic features. In our terms, irregularity is confined to the subset of syntactic features. This follows very straightforwardly from the model we have adopted: both the semantic and phonological lexicons must pick out elements stated in syntactic features (which are all that exist at syntactic levels of the grammar) and cannot diffferentiate between elements which are not syntactically differentiated.

This in turn implies that there is a finite (probably small) set of universally available semantic features which may be encoded syntactically in a given language. Such a hypothesis makes predictions both cross-linguistically and about language acquisition; we expect the child to learn syntactically-encodable semantic features in a different way from other, purely semantic features.

Now in Chapter 1, we suggested that we could overcome the "lexical learning" problem in language acquisition by adopting a Parallel Distributed Learning mechanism of the type advocated by McLelland and Rumelhart (1986), which made the right types of prediction about "lexical" learning (in particular, accounting for token variability in real-time language acquisition). We can now hook up the McLelland/Rumelhart learning model to the open/closed class distinction as encoded

by Emonds through the Designation Convention, and make a strong prediction about the type of learning which will characterize the auxiliary system: we should expect the same type of "U-shaped learning curve" to characterize the acquisition of auxiliaries in English as characterizes the acquisition of bound inflectional elements. We will return to this prediction at the end of the chapter, when we investigate the learning of the English auxiliary system.

Now, however, let us turn back to our starting point, the English auxiliary system, in order to see how the model we have developed accounts for the various phenomena discussed in 3.1.

3.5 Auxiliary Restrictions Revisited

3.50 Ordering Restrictions

Within our model, the (syntactic) ordering of G-heads and θ-heads is determined by the (syntactic) Canonical Government Direction (CGD) of the language in question (see 2.9). In English the syntactic CGD is rightward; therefore G-heads will precede their "complements". This will automatically account for the ordering of [I] (a G-head) and [V] (a θ-head), and thus for the ordering of modals and "raised" auxiliary verbs with respect to VP.

However, the CGD will not account for invariant ordering within VP. Thus the order of auxiliary have and be, for example, will not fall out from any ordering principle so far

proposed.

A number of possible ways of capturing these ordering relations suggest themselves. We could, for example, use morphophonological subcategorization as a filter on possible orderings. In the case of have and be, however, this would not work, since non-auxiliary have occurs - with appropriate morphology - in the proscribed order:

(76) Are we having fun yet ?

Alternatively, we could employ the semantics of auxiliary verbs to rule out the ungrammatical sequence (auxiliary be - auxiliary have). This would involve the claim that a sentence such as

(77) *We are having left.

was semantically anomalous, as opposed to

(78) We have been leaving.

I do not wish to explore the semantics of aspect in detail; nevertheless, a semantic approach seems promising, since, as first pointed out by Emonds (1976), "aspectual" verbs - those involving duration, iterative action, inception and completion - are generally incompatible with perfective have, as illustrated by (79):

(79)a. *I kept having spent too much on drink by the end of the month.

b. *I continued having run out of money before my cheque came.

c. *I started having left by the time my landlord arrived.

d. *I stopped having taken the beer out of the fridge when he did show up.

If we treat progressive be as an aspectual verb of the same type as those in (79), then we can account for its incompatibility with perfective have without further stipulation.

This leaves us with the ordering relation between progressive and passive be. Again, morphophonological subcategorization cannot account for the ordering facts, since the same (phonological) lexical item is involved in both possible orderings. However, this time aspectual restrictions cannot account for the ordering either; passive be is not an apectual auxiliary but part of an operation on thematic role assignment, and thus differs completely in semantic function from progressive be. Note also that there is no relation of "intrinsic closeness" between passive be and a passivized verb, since passive be raises into INFL (if no auxiliary intervenes) across an intervening not, just like aspectual be.

If we consider how passivization might operate on progressive (or main verb) be, however, we can come up with a plausible explanation for the ordering restriction. be, like such verbs as seem, become, and the aspectual verbs discussed

above, may assign only one θ-role. Now, passivization involves the suppression of a θ-role; it follows that if be were to undergo passivization, it would assign no θ-roles. But this situation is precluded in the theory adopted here; therefore be (and the other verbs mentioned above) cannot undergo passivization, which in turn means that they may not follow passive be.

We thus have three rather different explanations for the ordering restrictions holding between auxiliary elements, one syntactically-based, one based on aspect, and one based on θ-roles. Note, however, that each of these three receives support from similar co-occurrence restrictions holding elsewhere in the grammar; we have not needed to appeal to any properties unique to the auxiliary system. This constitutes a clear advantage in terms of learnability; the child needs to learn nothing specific to the auxiliary system in order to acquire its complex ordering restrictions. It also makes a prediction about the actual process of learning; we do not expect a child to make auxiliary ordering errors in the course of language acquisition, as long as the required distinctions have been learnt in the rest of the grammar. We will return to this prediction in 3.8, when we discuss the auxiliary system from the point of view of language acquisition.

3.51 Morphological Restrictions

We will have little to say about morphological restrictions. We

assume that in the morphophonological lexicon individual lexical items subcategorize for particular verb-forms, with a default value being assigned where no lexical subcategorization is required. Thus auxiliary have, for example, subcategorizes for a [V] with past-participle morphology, progressive be for [V] with -ing. It may well be that the relation between particular verb forms and the elements which subcategorize for them is non-arbitrary: see Fabb (1983) for a possible generalization.

3.52 SAI

We will not deal with SAI in any detail in this chapter; it will be the subject of an exhaustive examination in the next.

3.53 Negation

We have claimed that not is opaque for PF attachment of [V] to [I], without giving any real reason why. We are not sure there is any reason to give; it is certainly the case that this is a language-specific property of English (one shared by French, but not by Norwegian, for example), and an idiosyncratic property of not even in English (other negative elements do not have the same blocking effect). For want of any more explanatory account, we will simply assume that not is an opaque element for G-visibility at PF.

We will also assume that the normal position of not is

adjoined to VP; an optional process of cliticization attaches it to INFL, from where, of course, it can undergo SAI.

3.54 Floating Quantifiers

In 3.14 we showed that certain quantifiers can "float" to any pre-verbal position except that between passive or main verb be and its predicate, just in case progressive be is present. If we assume that these quantifiers may be adjoined to any [V] projection, this is accounted for, with the exception of the ungrammaticality of quantifiers in post-progressive pre-passive/main verb be position, as shown in (80):

(80)a. *We never realized that we could have been being all fooled, all the time.

b. We never realized that we could have been all fooled, all the time.

This type of exceptional behaviour also characterizes pre-passive post-progressive be in "VP-topicalization" contexts, which we will discuss in 3.57 below; we therefore postpone discussion of (80) until then.

3.55 Adverbs

Our treatment of not as VP-adjoined except when syntactically cliticized to an I-element has one advantage over an account where not is uniformly generated in INFL; it predicts the possibility

of an adverb intervening between an I-element in INFL and not. Sentence adverbials do indeed occur in such a position:

(81) John will probably not realize how much damage has been done.

On the other hand, if not is in INFL, then the adverb must be inside a lexical constituent, not a very attractive possibility.

3.56 VP-Rules

(i) "VP-deletion"

We will not go into details here about all the restrictions on this rule, some of which may not be syntactic at all; for contrasting approaches within GB theory, see Zagona (1982), Haik (1985), and Lobeck (1986). In our terms, the "deletion site" must be controlled by an [I] element. An additional condition will have to be added to deal with those varieties of American English which prevent gerunds or progressive be from controlling an empty VP:

(82) *Everyone else is being promoted , but Bill isn't being.

A possible explanation for this condition is given is provided by be-restructuring, which will be discussed immediately below.

(ii) "VP-fronting"

In general, "VP-fronting" (which we have claimed is better analyzed as a subcase of topicalization) fronts a "minimal" VP, excluding auxiliary elements. However, as we have already mentioned, be is sometimes fronted in VP-fronting environments:

(83) I said she would be a fool, and being a fool she is !

Notice that it is only main verb or passive be preceded by progressive be which behaves in this fashion; any other instance of fronted be is ungrammatical, as shown by the following examples:

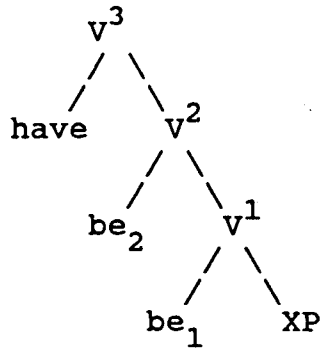
(84)a. *I said she could be a fool, and be a fool she can !

b. *I said she could be a fool, and been a fool she has !

This has led ASW to propose that, just in case progressive be is not present, main verb/passive be "restructures" into an auxiliary position, from which it may participate in processes such as SAI.

The process postulated by ASW involves the configuration in (85):

(85)



Here $\underline{be}_1$ is main verb/passive and $\underline{be}_2$ progressive <u>be</u>. The crucial point is that the V^2 projection serves as a target for the restructuring process; were it not present, there would be nowhere for restructuring to restructure into.

ASW's analysis poses potential problems for an approach such as our own. In particular, it implies that a V^2 projection is present at NP-structure (since restructuring "feeds" <u>have</u>-<u>be</u> raising). However, only aspectual <u>be</u> (or main verb/passive <u>be</u> after restructuring) ever occupies a position immediately dominated by V^2; and since aspectual <u>be</u> is in general optional, there is no independent motivation for assuming an empty aspectual position in V^2 to serve as the target for restructuring. The only thing that could ensure the presence of a V^2 projection would in other words be a phrase-structure rule which forced it to exist.

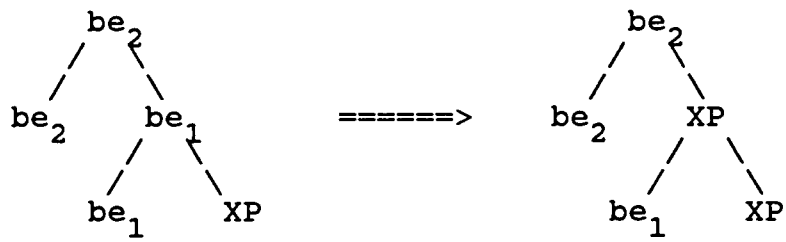
Within a model such as ours, this is an untenable proposition. Moreover, while ASW argue that there is independent motivation for a V^2 level, we saw in 3.22 that such motivation was flimsy, and more to the point, based on evidence (presence versus absence of auxiliary <u>have</u> in imperatives, the

why (not) construction, and complements to make and let) which is very unlikely to be accessible to the child. If so, we would have to postulate the V^2 level as not only universally available but unmarked (since it couldn't be easily learnt from Primary Linguistic Data).

Fortunately, however, there is an alternative which captures the main advantages of the ASW solution without its unpleasant consequences, and which, furthermore, results in a more plausible rule of restructuring.

Note first of all that the rule of be-restructuring restructures be out of its original position and into what appears to be a position adjoined to a higher projection. It thus represents a phrase-structure "expansion". It is this expansion which causes problems for our analysis. Suppose we said instead that main verb/passive be restructured into the minimal VP just in case it was governed by progressive be. In that case, restructuring would involve phrase-structure "contraction", as in (86) below:

(86)



Besides avoiding the problems mentioned above, this process has the added advantage of providing a potential explanation for the failure of be+ing to control a VP-deletion site (as

mentioned in 3.56 (i) above); if the restructuring process were obligatory, then be+ing could not be separated from the projection of [V] into which it was restructured, and consequently could not be "stranded" by VP-deletion. In addition, this analysis accounts for the quantifier floating facts discussed in 3.54, where a quantifier cannot intervene between be+ing and a following XP; we can simply state that the quantifier cannot appear inside a restructured constituent.

3.57 Imperatives

Imperatives are a problem for everyone We will assume that there is a highly defective imperative paradigm, and that none of the auxiliaries except the semantically neutral do have imperative "morphology". Note that in negative imperative contexts, even be and have require do-support:

(87)a. Don't be a bad boy !

b. ?Don't have let the chickens out by the time I get back !

Why (not) constructions are even more of a problem. We will assume they have a defective phrase structure, "missing out" I-projections and simply containing a VP with an adjoined WH-word. Do-support is not triggered, since there is no tense present for not to interrupt.

3.58 Complementation

(i) Infinitives

Our account of infinitives assumes that to is a [-TENSE] [I] constituent, which contains no AGR-specification; in that case, it is not subject to AGR-visibility, and need not be associated at PF with [V]. This accounts for its comparative syntactic freedom, and for the absence of modals and do-support in infinitivals, since the former are inherently [+TENSE] and the latter is only inserted to satisfy AGR-visibility.

(ii) Gerunds and other "defective complements"

Non-nominal gerunds are instances of VP-complementation. This immediately provides an explanation for most of the auxiliary restrictions associated with "poss-ing" complements: modals are not generated in VP, therefore never appear in gerunds; there is no tense, therefore negation will never trigger do-support. Note, however, we will still need some version of the double-ing constraint to prevent progressive be from occurring in gerunds.

Our analysis of let and make is very similar, except that these verbs will subcategorize for a bare stem rather than an -ing-affixed V.

Turning to perception verbs and verbs of temporal aspect, we again claim that these are instances of VP-complementation. We will remain vague about the restrictions on auxiliary have and

be in the complements of these verbs (see 3.18 above), partly because judgements are variable and partly because we suspect that semantic rather than syntactic factors determine their distribution, though, as mentioned, no semantic account quite fits all the facts. We will have to leave this question open.

This concludes our analysis of the English auxiliary system. I hope to have shown that though many residual problems remain, our approach at least promises an empirically adequate and explanatorily illuminating account of one of the most difficult areas of English syntax. In the next sections, we will move onto a (limited) cross-linguistic look at auxiliary systems, with special reference to issues of learnability.

3.6 The Auxiliary System Cross-Linguistically

3.60 Introduction

In the following sections, we will extend our analysis of the English auxiliary system to a (limited) set of cross-linguistic data from Germanic and Romance languages. It will be shown that the system outlined here generalizes to I-elements in other languages without further stipulation; in fact, cross-linguistically, it is the English system which turns out to be marked. Finally, in 3.7, we will attempt an explicit parametrization of the system outlined here, before going on in 3.8 to take a look at the acquisition of some auxiliary processes.

3.61 INFL in German

Whereas in English there is solid evidence for at least a three-way partition in tense-bearing elements (which we captured through the use of three levels of [I]/[V] association) no such evidence exists in German (or any other Germanic language). On the contrary, "auxiliary" verbs in German behave like other verbs, as shown by Steele et al. (1981), from which the following examples are taken. First of all, modal and other auxiliary verbs display full inflectional paradigms, as illustrated by the verbal paradigm for <u>können</u>, "to be able":

(88)	Present	Past
1st sing.	kaan	konnte
2nd	kaanst	konntest
3rd	kann	konnte
Ist pl.	können	konnten
2nd	könnt	konntet
3rd	können	konnten

Of course, this in itself is not enough to decide the status of auxiliary elements in German; as we have seen, the modal <u>do</u> in English has a full-blown verbal morphology even though it never appears in VP. More to the point, there are no German equivalents of the other, inflectionally defective, English modals; thus there will never be any morphological evidence for a child learning German that there exists a separate class of modal elements. Turning to other types of evidence, Steele et

al. note that German modals, unlike their English equivalents, occur in non-finite clauses:

(89) Er scheint gehen zu können.
He seems go to can

"He seems able to go."

Unlike English modals, German auxilaries may cooccur, in any order (though obviously, changing the order changes the meaning of the resulting sentence):

(90)a. ...weil Fritz kommen wollen konnte.
...because come want could

"...because Fritz could have wanted to come."

b. ...weil Fritz kommen können wollte.
because Fritz come can wanted

"...because Fritz wanted to be able to come."

German modal auxiliaries can occur in certain circumstances as main verbs, in contrast to their English equivalents:

(91)a. Hans kann keinen Handstand.
Hans can no handstand

"Hans can't do a handstand."

b. Ich muss nach Hause.
I must to home

"I must go home."

Finally, whereas "contractability" acts as a diagnostic test for auxiliary elements in English, equivalent elements in German never contract:

(92)a. Er will gehen.
He wants go

"He wants to go."

b. *Er'll gehen.

(88 - 92) show that there is no lexical [I,V] element in an [I] position in German; the same holds true, as far as I know, for all other Germanic languages. Some authors have used this to argue that there is no [I] element at all in German; we are unable to accept this conclusion both for conceptual reasons ([I] is, we claim, a universal category-feature) and empirical ones (without [I], it is impossible to propose a general motivation for SAI-type rules, which is one of the main purposes of this thesis). Moreover, since nominative Case-assignment only takes place under (maximal) government by [I], the GREPP will ensure the presence of an I-element at NP-structure.

Given, then, that in German no [I,V] element occupies an [I] position, yet an [I] element must be present, it remains to ascertain how [I]/[V] association takes place. In our investigation of the English auxiliary system, we saw that association could take place at any one of three levels: in the lexicon (for modals), in the syntax (for have/be) or at PF (for ordinary verbs). The question now arises as to which, if any, of these three options adequately characterizes the German [I]/[V] system.

In fact, I will argue that German does not employ any of

them; instead, I will propose a fourth option, where [I,V] constituents are generated directly "in VP". This possibility will in turn be related to other differences between German and English syntax, relating to NP-movement, Case-assignment, and the ECP; it will be shown that several independent phenomena receive a unified explanation if [I,V] elements are present in VP. This means that no independent [I] node exists in the syntax in German, and that "V2" movement takes place directly from VP to "COMP", without passing through an intermediate INFL node; the implication is that "V-raising to INFL" and "V-raising to COMP" are separate processes which may occur independently of each other.

It is easy to show that [I,V] association must have taken place prior to "V2" movement, which we will take to be a syntactic process (i.e., applying in the mapping from NP-structure to S-structure). Any inflected main verb can appear in inverted position, as in yes/no questions:

(93)a. Sprichst du Englisch ?
Speak you English

"Do you speak English ?"

b. Hast du Englisch gesprochen ?
Have you English spoken

"Did you speak English ?"

Note that there is no difference between the main verb sprechen and the auxiliary haben with respect to inversion; both may invert freely as long as they bear the relevant tense-features.

Since only [I,V] elements may invert, it follows that [I]/[V] asssociation must have applied prior to inversion.

Given then that there must be at least some [I]/[V] association prior to PF, it remains to be seen whether Germanic also employs PF-association. One possible test is provided by negation: if negative particles behave in German in the same way as in English, then we should expect V-raising verbs to move into INFL over the negative particle. If, as commonly assumed, INFL follows VP (German being partially left-branching) this means we should expect the negative particle to follow a non-raised verb but precede a raised verb. Unfortunately, things aren't that simple: though a negative particle in German may never intervene between an auxiliary and a main verb, as shown in (94b.), in fact in general the negative particle nicht precedes both main verb and auxiliary, as shown by (94a.) and the ungrammatical (94c.):

(94)a. Ich sagte daβ ich den mann nicht gesehen habe.
I said that I the man not seen have

"I said that I didn't see the man."

b. *Ich sagte daβ ich den mann gesehen nicht habe.

c. *Ich sagte daβ ich den mann gesehen habe nicht.

(Note that embedded clauses are used to avoid the distorting effect of the V2 condition).

The data in (94) could be subjected to various interpretations. We could simply assume that the position of the negative particle is subject to cross-linguistic variation, being VP-

adjoined in English but V-adjoined in German. Aside from its lack of explanatory adequacy, however, this account runs into serious empirical problems. Firstly, it would have to be stipulated that nicht left-adjoined to the first in a string of verbs, in order to avoid generating (94b.); such a structure-independent condition would be hard to justify. Secondly, we would predict that in V2 clauses nicht would front along with a tensed verb: this prediction is false, as shown by (95):

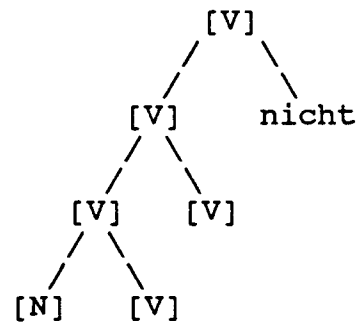
(95)a. Ich sah den mann nicht.
I saw the man not

b. *Ich nicht sah den mann.

c. *Ich sah nicht den mann.

Let us assume, then, that nicht is VP-adjoined, as in English. In that case, the underlying structure of the embedded VP in (94a.) would look something like (96):

(96)

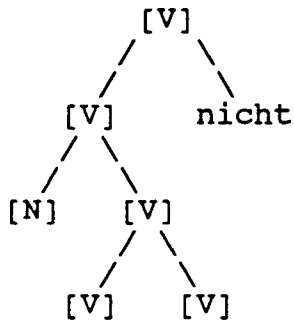


Now, in order to generate a grammatical surface-string, some kind of [V] movement must take place. There are two possibilities for such a rule: it could either involve the lexical head [V], or some projection of [V].

Suppose it were a case of head-movement. Clearly, it could not be a case of V-raising to INFL, since such a rule crucially refers to finite verbs, and the process in question raises all of the verbs in a string out of VP. In fact, any adjunction of a lexical [V] to a phrasal category would be independently ruled out by the Head Movement Constraint of Travis (1984), which prevents lexical heads from moving anywhere except into a governing head position.

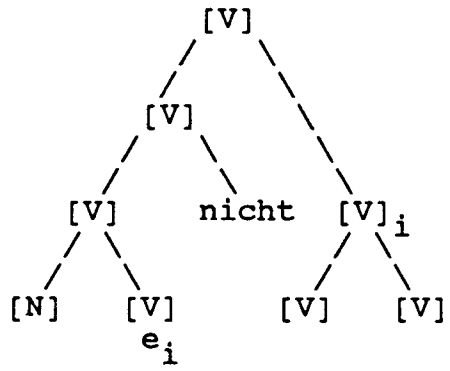
Suppose, then, it were a case of [V]-projection movement. Such an analysis would also be problematic, since the relevant projection would include a string of verbs but no complements - in other words, a non-constituent. The obvious move to make here would be to postulate a restructuring process which would turn a structure such as (96) into one such as (97):

(97)



Following restructuring, [V]-projection raising would apply to create the structure in (98):

(98)



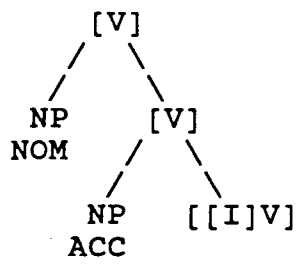
Interestingly enough, independent work on "V-raising" in German and Dutch (see Evers 1975) has led Haegeman and van Riemsdijk (1986) to postulate that an almost identical process, involving restructuring and verb-projection raising, takes place in infinitival clauses in German and in Dutch. This suggests that the two might be incorporated, and that the range of V-projection movement in Germanic is greater than previously imagined. We will not pursue the many interesting questions raised by such a possibility, since however we treat V-projection raising, it clearly has no bearing on the question of whether PF [V]/[I] association is to be permitted.

In fact, we can ask a rather more fundamental question: given the V-projection rule outlined above, is there any justification for the assumption that German also employs V-raising to INFL ?

I would like to claim that there is not. Instead, I will propose that in fact, unlike English, German allows [I,V] categories to be "base-generated" in VP; more specifically, I propose the following NP-structure representation for an

embedded clause such as that in (94a.):

(99)



In other words, German has lexical [I,V] association, but in precisely the opposite sense to that of English; whereas in English lexical [I,V] constructions a verbal stem is "absorbed" by an inflectional element, in German the inflectional element is absorbed by a verb.

Certain additional assumptions are necessary in order to make (99) work; in particular, it must be assumed that the [V] constituent may "adopt" the nominative Case feature attached to [I] (a suggestion independently motivated by our discussion of the ergative Case-marking parameter in 2.4), and we need a principle of Case-conflict (provided in (100) below).

The structure in (99) is motivated by several properties of German syntax. First of all, there appears to be good evidence that nominative Case can be assigned VP-internally under certain circumstances. This evidence is provided by the phenomenon of "Nominative/Dative Inversion" (den Besten 1985, Webelhuth 1986b.). The unmarked order of NPs in the "Mittelfeld" in declarative clauses is nominative-dative-accusative; however, in passive and "unaccusative" structures (see Burzio 1981) the unmarked order is dative-nominative. This

is easily explained if it is assumed that no NP-movement takes place in German; instead, the internal argument of passivized and unaccusative verbs simply takes on the external Case, i.e., nominative. This hypothesis involves two other premises: firstly, nominative Case is obligatorily assigned (just as in other nominative/accusative languages); and secondly, nominative Case may be assigned VP-internally. Webelhuth (1986b.) gives several supporting arguments for this second assumption, based on the distribution of definiteness effects and the constituency of the [NP V] string in unaccusative and intransitive sentences; we will not review these arguments here.

Unlike Tappe (1982), Haider (1985, 1986) and Sternefeld (1985) we do not assume that German has a "flat" structure, (i.e., without a syntactic VP) in spite of the fact that nominative Case may be assigned in VP. German shows at least some configurational properties: in particular, it shows weak crossover effects (Webelhuth 1986a.) and a partial ECP asymmetry based on sub-extraction of a WH-phrase from the <u>was für</u> construction (Bennis 1983, den Besten 1985). These "configurational" properties are independent of surface Case-marking (for example, den Besten shows that nominative NPs in subject position and in direct object position behave differently with respect to <u>was für</u> extraction), apparently refuting the claim that there is no need in German for a notion of subject distinct from that of nominative Case-marked NP.

Accordingly, the structure in (99) allows nominative to be

assigned in VP, by generating [I,V] elements in VP, while maintaining a configurationally defined distinction between subjects and non-subjects.

It might be asked exactly how Cases are assigned in a configuration such as (99). Following Czepluch (1983) and Haider (1985) we will assume a distinction between structural and lexical Case. The former includes the Cases represented by the nominative/accusative and ergative/absolutive Case-systems, the latter the "oblique", lexically specified Cases, including Dative and Genitive in German, for example. Various criteria can be adduced in support of this distinction, including "quirky" Case, adjacency, lexical idiosyncracy, and optionality: see Haider (1985) for the relevant arguments.

Modifying ideas of Czepluch (1983), we will assume the following principle:

(100) Principle of Case-conflict:

> Structural Cases must be assigned at different levels of phrase structure.

Principle (100) prevents any language with a structural Case-system from having a "flat" structure (though it leaves open the possibility that languages which only have lexical Case will be "non-configurational" in this sense). Intuitively, (100) is a generalization of the notion of "Case-conflict" which prevents an argument from receiving two Cases. In

German, which has a nominative/accusative structural Case-system, (100) will force nominative Case-assignment to take place on a different structural level than accusative. Now, since in nominative/accusative languages accusative Case is normally assigned to the internal argument of the verb, nominative (the external Case) will have to be assigned at a higher structural level. Let us assume (following Fabb 1984, Massam 1985 and much other work on Case-theory) that Case-assigning categories are marked with the feature [+Case]. A "derived" [I,[V]] category containing a Case-marking [V] will then have two Case-features, as in (101):

(101) $[I_{[+Case]}, [V_{[+Case]}]]$

Now, in transitive sentences, the [V] constituent will be able to assign its Case-feature to an internal argument, leaving nominative to be assigned to the external argument, which, by (101), will be realized at a different phrase-structure level. In the passive/unaccusative cases, accusative Case will be suppressed; in that case, nominative can be assigned to the internal argument of VP without violating (101), thus accounting for nominative Case-assignment "inside" VP.

The question arises as to the status of the subject position in the latter type of structure: given that nominative Case is assigned to an internal argument, and predication is also satisfied by this argument (which minimally governs AGR), is there any justification for a separate subject position ?

We will assume there is not. This means that in many ways German finite clauses more closely resemble "small clauses" than English-type tensed clauses. It is beyond the scope of this chapter to pursue the many implications of this view of German clause-structure.

Let us instead return to our more limited original aim of accounting for [V]/[I] association.

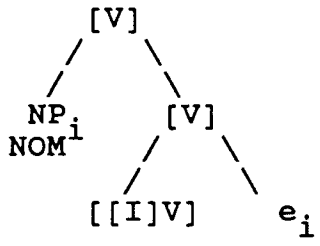
We have assumed that in German [I,V] constituents occupy [V] positions, whereas we have argued that in English [I] elements only occur in [I] positions, necessitating [V] raising to INFL at some level of the grammar. This correlates with the ability of [V] in German to absorb [I], whereas in English, [I] absorbs [V]. In this sense, English is "I-strong", whereas German is "V-strong".

It must now be asked if this difference can be reduced to more primitive (and thus ultimately more learnable) differences between the two languages.

As a first observation, notice that in English external and internal Case-assignment occur in different directions, whereas in German they occur (according to the view advocated here) in the same direction. If we were to hypothesize that a single constituent could only assign Case in one direction, then we could account for the failure of [I,V] constituents to occur at NP-structure in English except in the limited case of modals, which assign no Case. We would not account for passives and unaccusatives in English, however, which also assign no internal Case yet still cannot be generated "in VP" as [I,V]

constituents. Suppose an [I,V] constituent were to be generated "inside VP" in a passive/unaccusative structure in English, as in German. In that case, we would get a configuration such as that in (102):

(102)



It is difficult to see how to avoid generating such a structure in English, since it violates none of the syntactic principles we have proposed so far. Note, however, that structures such as (102) involve Case- and θ-role assignment in different directions. We might then propose the supplementary condition (103:

(103) Directional Uniformity:

A single constituent must assign Case and/or θ-roles unidirectionally.

(103) will successfully differentiate German passives and unaccusatives from those in English, though its status as a principle of UG is somewhat dubious; in particular, if there are separate parameters governing direction of Case- and θ-role assignment, as claimed by Koopman (1984) and Travis 1984),

Directional Uniformity will have to be recast as a markedness statement, rather than an exceptionless principle. Note also that (103) resembles the θ-Case correspondance principle introduced in 2.8 as a learning heuristic for phrase-structure; it may be that Directional Uniformityis not a primitive of the theory, but may be derived from the interaction of more fundamental principles with learnability considerations. We will not, however, attempt to investigate further here, leaving (103) as a stipulation that we hope will ultimately fall out from more general principles.

To conclude our discussion of German (and, incidentally, Dutch, which acts identically in the relevant respects), let us summarize the main differences between German-type [I]/[V] association and English-type [I]/[V] association. The three relevant properties are schematized below:

(104)

	English	German
Nominative in VP	-	+
NP Movement	+	-
V-raising to INFL	+	-

We have argued that all three properties relate to the ability of German to generate [I,V] constituents "in VP". Furthermore, we have suggested that the different [I]/[V] association mechanisms of English and German might be reducible to even more basic principles involving uniformity of directionality in

Case- and θ-role assignment, principles which we have argued (in Chapter 2) meet the criterion of epistemological priority; thus we hope to have laid the basis for an explanatory account of the inflectional parameters which distinguish the two languages.

3.62 INFL in Scandinavian

The Scandinavian languages are particularly interesting from our viewpoint in that in some ways they closely resemble German and Dutch (sharing the V2 condition, for example) yet in others they more closely resemble English. In particular, all have a basic SVO order (as opposed to the SOV order of the W.Germanic languages) and therefore, like English, assign nominative and accusative Cases in different directions. This gives us the opportunity to test out our hypotheses concerning directional uniformity and [I]/[V] association: we should expect the Scandinavian languages to resemble English in not permitting [I,V] constituents to be generated in VP. This in turn means that we should expect such languages to share a cluster of INFL-related properties with English: they should exhibit overt NP-movement, V-raising to INFL, and fail to assign nominative Case in VP.

We will illustrate the relevant properties with examples from Norwegian; the other continental Scandinavian languages behave much alike.

First of all, let us examine NP-movement. In a language such

as English, NP-movement is marked at NP-structure by the presence of an empty category in internal argument position, which is A-bound by an NP in subject position; on the other hand, in a language such as German, there is no empty category, but simply a switch in Case-assignment possibilities, together with the suppression of the θ-role normally assigned to subject position. If Norwegian was to pattern like German, we should then expect passive sentences, for example, to show a $[_{VP}$ V $NP_{NOM}]$ configuration. However, in fact, Norwegian passives are much like those in English, as illustrated by (105):

(105)a. Bill så John.
"Bill saw John."

b. John ble sett.
was seen
"John was seen."

Thus the first of our predictions is borne out.

Next we turn to V-raising to INFL. Suppose we use the relative position of the inflected verb and the negative particle as a diagnostic for V-raising. If Norwegian resembles English, then V-raising should lead to an order of $[V/INFL_i$ NEG $[_{VP}$ e_i ...]] in negated sentences. However, this turns out not to be the case; once we have abstracted away from the "V2" condition, tensed verbs in Norwegian <u>follow</u> both the negative particle <u>ikke</u> and other sentence adverbials, as shown in (106):

(106)a. Jeg sa at Bill ikke så John.

I said that not saw

b. *Jeg sa at Bill så ikke John.

"I said that Bill didn't see John."

c. Jeg sa at Bill ikke hadde sett John.
not had seen

d. *Jeg sa at Bill hadde ikke sett John.
had not seen

"I said that Bill hadn't seen John."

In the face of this evidence, we can make one of two moves: abandon the V-raising hypothesis, or abandon our V-raising diagnostic of "negative opacity". By choosing the former option, we lose a whole set of similarities between Norwegian and English; by choosing the latter, we lose nothing; in fact, we characterize the behaviour of <u>not</u> as exceptional, which, as noted above, is true even in English.

This means that, since there are no "opaque elements" for PF attachment, PF [I]/[V] association will represent the normal state of affairs in Norwegian; V-raising, on the other hand, will take place just in case the "V2" condition applies (in conformity with Chomsky's "last resort" theory of movement rules). THis conclusion conforms with learnability considerations: if PF-association was marked relative to syntactic V-raising, then the child learning English would need to note the failure of main verbs to raise into INFL in negative sentences or to invert in SAI contexts in order to learn that PF-association took place in English. The availability of such evidence does not provide insuperable

learnability difficulties, but the complete absence of overgeneralizations involving V-raising in the acquisition of the English auxiliary system (see section 3.8 below) does argue that PF-association is not a particularly marked option.

Thus we see that Norwegian bears out our predictions concerning directional uniformity, behaving much like English and quite unlike German and Dutch in the relevant respects. It differs from English, however, in several other ways: it allows any tensed verb to occupy the "V2" position in inversion contexts - and, by implication in V-raising contexts; it treats the negative particle as "transparent" for PF-association; and it does not allow lexical [I,V] attachment in [I], as exemplified by the English modals. It will be noted that in all these respects English constitutes a "marked" system; as such, we might expect differences between the learning of the English and Norwegian auxiliary systems. While, unfortunately, comparative data is not available, we shall see that the "marked" character of the English system is indeed reflected in acquisition data.

3.63 INFL in Romance Languages

So far, we have provided evidence for the existence of two rather different types of inflectional system. On the one hand, there is the Scandinavian/English type system, which is characterized by V-raising to INFL and NP-movement; on the

other hand, the German/Dutch type system, characterized by neither of these properties. We have also shown that this contrast is independent of V-to-COMP rules such as the "V2" condition and SAI; here West Germanic languages pattern with Scandinavian languages, in contrast to English.

We now turn to a third and rather different set of languages, including Italian, Spanish, Catalan, Portugese, and Rumanian (but excluding French) which are characterized by the following properties:

(a) They are all (with the possible exception of Portugese) subject to what we shall call (following Safir 1985) the Free Inversion Parameter (FIP) - see Chomsky (1981), Jaeggli (1982), Rizzi (1982), Safir (1985), Bouchard (1984), Borer (1986), amongst others. Languages with Free Inversion may generate subjects in either preverbal position or in a post-verbal position generally analyzed as VP-adjoined (see Rizzi 1982 for justification). Thus in standard Italian, for example, either (107a.) or (107b.) is grammatical (in fact, the latter is rather more common):

(107)a. Gianni ha telefonato.
Gianni has telephoned

b. Ha telefonato Gianni.
Has telephoned Gianni.

"Gianni has telephoned".

Languages subject to the FIP typically show no complementizer-

trace ECP effects; following Rizzi (1982) it is widely assumed that the post-verbal subject position is lexically governed by the verb.

(b) Most, but not all are subject to what we shall term (again following Rizzi 1982) the Null Subject Parameter (NSP). Thus in tensed clauses in Spanish, most dialects of Italian, Portugese, and Catalan (but not in Trentino or Romanian) phonetically realized subjects are optional, whereas in English they are not. Thus in Italian, (108) is quite grammatical, and in fact overt pronouns are usualy only used for emphasis or clarification:

(108) Ha telefonato.
3-S has telephoned

"He has telephoned."

The most common explanation for the cross-linguistic variation associated with the NSP is the theory of "rich" AGReement first suggested by Taraldsen (1979) and subsequently explored in numerous publications, including Rizzi (1982, 1986), Chomsky (1981, 1982), Borer (1984) Jaeggli (1982), Safir (1985), and Roberge (1986). The basic idea behind all these approaches is that null subjects may be generated where AGR is sufficiently "rich" to recover the syntactic features of an empty pronominal, variously described as PRO (Chomsky 1981, Jaeggli 1982), pro (Chomsky 1982, Roberge 1986) and "Expletive e" (Safir 1985). Thus in (108) above the inflectional morphology

of the auxiliary ha includes information on number and person which can "identify" an empty NP in subject position. Note that such information is also present in its English equivalent, has, yet no "pro-drop" is possible; obviously, more remains to be said. In fact, some languages, for example Hebrew, as described by Borer (1984), and Irish, as in McCloskey and Hale (1984), do show a "partial pro-drop" paradigm, allowing a null subject only where inflectional morphology allows syntactic recovery; for relevant discussion, see Roberge (1986)).

There has been considerable debate as to whether the NSP should be parametrically linked to the FIP; we will assume, with Safir and Roberge, that the two properties are independent. We shall have relatively little to say about the NSP here; the FIP, on the other hand, is obviously relevant to our concerns, in particular in so far as it relates to the behaviour of INFL. Accordingly, we will briefly discuss the main approaches taken within GB theory to the FIP, before going on to propose our own variant.

At the heart of what we might call the standard GB approach to the FIP is the process named "Rule R" by Chomsky (1981). In fact, Rule R is none other than the familiar rule of Affix Hopping discussed above (and ad nauseum in introductory syntax classes). Rule R "lowers" INFLectional features onto the verb in VP (thus constituting the standard GB equivalent to our I/V association processes). In English, this is a PF process whereas in FIP languages, it is hypothesized, Rule R takes place in the syntax. This is important, because syntactic

affix-hopping allows nominative Case to be assigned within VP (to a post-verbal subject), thus accounting for the ability of a nominative NP to show up in VP in a FIP language such as Spanish but not in English.

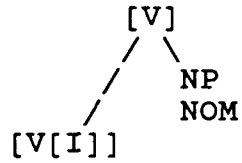
Now, this analysis begs as many questions as it answers. In particular, the movement of the subject from an [NP,S] position to an [NP,VP] position violates well-motivated constraints on movement rules, since it is a "lowering" rule. In particular, it runs into trouble with Binding Condition C, since the post-verbal NP is coindexed with a c-commanding antecedent, even though it is an R-expression and should be free. This problem cannot be simply solved by denying that the two positions form a chain, since under standard assumptions a (non-ergative) post-verbal NP must inherit its θ-role from the [NP,S] position. Various proposals have been made to deal with this contradiction, including the introduction of a special "co-superscripting" relation between INFL and the post-verbal NP (in Chomsky 1981) and the notion of an "external θ-set" including all NP sisters of VP (in Safir 1985).

Another set of problems concern Rule R. So far, we have seen the operation of a syntactic raising rule (V-raising) and a PF lowering rule ("affix hopping"); Italian, however, seems to require a syntactic lowering rule. Now, under a plausible interpretation of the "Head Movement Constraint" (Travis 1984) the fact that the <u>only</u> lowering rules allowed seem to be local PF processes falls out from the ECP, which as an LF licensing condition either applies at LF or at S-structure, or probably

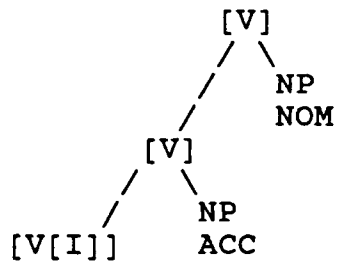
both, and requires a c-commanding antecedent for cases of antecedent government, including head-movement rules. However, if the ECP applies to movement rules involving [I]/[V] association, then it should disallow syntactic lowering, and in particular, a syntactic version of "Rule R".

In view of these considerations, let us suppose that in fact neither movement to a post-verbal subject position nor Rule R exists. Instead, we will adopt an analysis similar to the one we have proposed for German: INFL in FIP languages may be base-generated in VP, which in turn allows for the possibility of a nominative NP in VP. Now, the principles of Case-conflict (100) and Directional Uniformity (102) predict the following structural representation for a sentence such as (107b.):

(109)



It will be observed that (109) is the mirror image of the structure suggested for German in (99) above; however, there is one respect in which Italian acts in a very different way from German. We might expect, given the type of structure in (109), that Italian should also allow structures such as that in (110):



But this prediction is false: in general, Italian allows only one (non-focussed) NP to the right of the verb:

(111)a. Ha mangia la mela
Has eaten the apple

"He has eaten the apple."

b. Ha mangia Giovanni

"Giovanni has eaten".

c. Giovanni ha mangia la mela.

d. *Ha mangia la mela Giovanni.

In order to account for this difference, we will make the following assumptions:

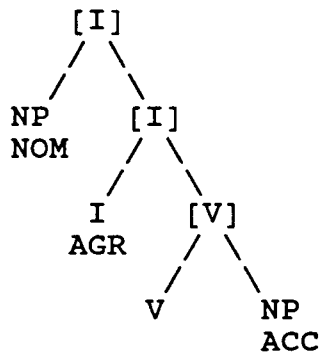
(a) V may only assign one Case in VP.
(b) V may only assign Case rightward.

(a) is possibly to do with Case-adjacency, though we will not speculate further; (b) is a consequence of directional uniformity, as encoded by (103) above.

Now, these two assumptions, together with our account of Case-marking in [I,V] constituents in VP, allows us to account

for the configurational properties of Italian. First of all, take a sentence such as (111c.). Here, [I,V] association resembles PF attachment in English, and thus we are justified in postulating a structure such as (112):

(112)



The same structure will suffice for NSP-structures such as (111a.), where AGR is "rich" enough to identify pro in subject position.

Turning to FI structures such as (107b.), (111b.), we will need to postulate lexical [I,V] association in VP, resulting in structures such as (109). Since [V] may absorb features of [I], including Case-features, it may assign Nominative Case; however, by (b) above, it may only assign one (structural) Case.

According to this account, then, we do not need to derive FI structures from structures such as (112); the two sets of representations result from two different possibilities of [I,V] association instantiated in the same language. This in turn follows from our conception of phrase-structure as itself essentially derivative from other principles; since we are not

committed to a single underlying PS representation, we do not have to encode the relation between SV and VS structures in Italian, for example, through a derivational history. Of course, this does not mean we are not committed to a restrictive theory of phrase-structure; but such a theory will fall out from limited parametric variation in other components, in this case from the different ways of meeting the AGR-visibility condition.

If, as indicated here, Italian allows both lexical [I,V] association in VP and PF association, then we predict that it should show both German-type behaviour and English-type behaviour with respect to NP-movement and V-raising. The latter provides no relevant evidence, since the negative particle <u>non</u> in Italian (and its equivalent in other Romance languages) is transparent to PF association, as shown in (113) below:

(113)a. Non ha telefonato Giovanni.
Not has telephoned Giovanni.

b. *Ha non telefonato Giovanni.

"Giovanni hasn't phoned."

On the other hand, NP movement does seem to show both English-type and German-type properties: for example, arguments of "ergative" verbs may remain in internal argument position, but they may also "raise to subject", leading to the alternative surface realizations in (114a.) and (114b.):

(114)a. Arriva Giovanni
arrives G.

b. Giovanni arriva
G. arrives

"Giovanni arrives."

Now, Burzio (1981) provides a good deal of evidence that in Italian the single argument of "ergative" verbs and of passive verbs (as in (114b.)) differs syntactically from the "inverted" subject of other transitive and intransitive verbs, and that the former should be base-generated in internal argument positions, whereas the latter occupy external argument positions. The relevant evidence, which we will not review in detail, includes the following syntactic tests: the behaviour of the clitic ne (which may form a clitic-chain with respect to all and only internal arguments); auxiliary selection (ergatives and passives take essere, transitives and intransitives avere); various linear order phenomena; sentential pronominalization; and "clitic climbing" in causative constructions. Burzio's conclusions indicate that, as predicted, Italian shows ambiguous behaviour with respect to NP movement; in passivized and unaccusative structures, the internal argument may either raise to subject, as in English or Norwegian, with a structure such as (112), or it may remain in situ, as in German, with the structure in (109).

3.64 INFL in French

French differs from Italian and Spanish in failing to allow Free Inversion. We might then expect it to pattern with English and the Scandinavian languages, in which case it should show a now familiar cluster of properties, including NP-movement, and V-raising to INFL. And indeed, French does show these properties. NP-movement works much as in English: passive, for example, involves raising of an internal argument to an external argument position:

(115)a. Un terroriste a tué le premier-ministre.
A terrorist has killed the prime-minister

"A terrorist killed the prime-minister."

b. Le premier-ministre a été tué par un terroriste.
The prime-minister has been killed by a terrorist

"The prime-minister was killed by a terrorist."

That French has V-raising to INFL is shown by the relative order of an inflected verb and the negative particle pas:

(116)a. Les autorités n'ont pas arrêté l'assassin.
The authorities have not arrested the assasin.

b. *Les autorités ne pas ont arrêté l'assassin.

c. *Les autorités n'ont arrêté pas l'assassin.

"The authorities didn't arrest the assassin."

(The scope-marker ne, marginal in colloquial French, has no bearing on the present analysis).

(116) shows that the negative particle pas is opaque to PF-association, triggering V-raising to INFL, as in English. However, in French, as opposed to English and Norwegian, V-

raising appears to occur universally, as argued by Emonds (1978). Thus both quantifiers such as tous in (117a.) and adverbs such as jamais in (117c.) appear in French after the tensed verb, rather than before, as in English or Norwegian:

(117)a. Les anglais détestent tous les cuisses de grenouille.
The English hate all the thighs of frog

b. *Les anglais tous détestent les cuisses de grenouille.

"The English all hate frogs' legs".

c. Quandmême, ils mangent souvent des choses moins
All the same, they eat often things less

délicieuses.
delicious

d. *Quandmême, ils souvent mangent des choses moins délicieuses.

"All the same, they often eat less tasty things".

Thus it appears that in French V-raising is obligatory, unlike in English, where it is obligatory only when not is present or when SAI applies, and unlike in Norwegian, where V-raising is only obligatory in order to feed V2.

Note that French resembles Norwegian and differes from English in its inflectional system, allowing any tensed verb to raise to INFL (whereas English only has the much more restricted rule of have/be raising).

3.7 Parametrizing the Model

3.70 Introduction

It is now time to take stock of the variation we have observed in our brief and obviously incomplete cross-linguistic survey, with a view to proposing an explicit parametrization of the auxiliary system. Such a parametrization will clearly be insufficient to explain many auxiliary phenomena in languages which we have not discussed, since it is beyond the scope of this thesis to attempt to account for all possible auxiliary systems. We will thus be forced to make quasi-universal claims on the basis of a small subset of available data, a position not unfamiliar to practitioners of generative syntax committed to the metatheoretical hypotheses of Universal Grammar but unable to provide the necessary empirical basis for their claims. It follows that what follows is indubitably wrong; however, as long as this is conceived of as the inevitable - if frustrating - fate of any particular episode in the Generative Enterprise, the enterprise itself need not be abandoned.

Let us then review the universal claims which we have made and the types of variation which we have observed.

3.71 INFL as a Universal Category

First of all, we have made the claim that the category [I] is universally present. Such a claim resembles in many ways that put forward by Steele et al. (1981) in a rather different framework, and for rather different reasons. Thus, whereas we have relied heavily on theory-internal considerations to force

the existence of INFL, or more precisely, AGR (its presence is necessary to satisfy P-licensing conditions on interpretation of "visible" elements at LF), Steele et al. take pains to produce - as far as this is possible - a relatively theory-independent definition of INFL (or AUX, in their terminology). Their definition involves

"...a specified (i.e., fixed and small) set of elements, crucially containing elements marking tense and/or modality..." (Steele et al., p.21)

Now, such a definition is meant as a kind of yardstick for establishing cross-linguistic equivalence classes, and as such combines criteria based on semantic properties with more strictly syntactic constituency tests. Nevertheless, the claim that a syntactic constituent is systematically associated with a particular set of semantic properties cross-linguistically is of considerable importance to our own hypothesis, since it provides a "semantic bootstrapping" mechanism for the child learning the syntactic constituent INFL/AUX. Steele et al. also identify several other properties of AUX constituents instantiated by a majority of the languages in their study: thus Luiseno, Lummi, and Egyptian Arabic all have some form of subject-marking encoded in the AUX constituent, a purely syntactic generalization which reinforces the conclusion reached by much recent GB work that a "special relationship" exists between INFL and the external argument of a predicate which it modifies. Other "notional categories"

commonly encoded by AUX constituents relate to non-declarative sentence-types, including negative, interrogative, quotative, and evidential sentence modalities.

The work of Steele et al., then, appears to independently support our conclusion that the category [I] is universally present in natural languages, and may be identified by a first-language learner on the basis of a consistent cross-linguistic syntax-semantics correspondance (see Kaisse 1985, though, for a different view).

3.72 Parameters of [I]/[V] Association

Given, then, that the category [I] is universally available, and assuming the universality of the P-visibility conditions and the AGR-visibility condition, the question of cross-linguistic variation comes down to variation in how [I]/[V] association is brought about. In this section we will summarize the mechanisms we have proposed, and suggest a parametrization.

In our discussion of English, we have identified three types of association, corresponding to three different grammatical levels. The first, lexical association, accounts for the characteristic properties of modal auxiliaries, which we have argued are inserted directly into an I-position outside VP. The second, syntactic association, involvs the "movement" of <u>have</u> and <u>be</u> into this VP-external I-position in the syntax. (Though in the model adopted here no actual movement takes place, a verbal trace exists at NP-structure "as if" it had.) The third,

PF association, involves a PF lowering rule which realizes I-features phonologically on the verb.

Extending our analysis to a limited cross-linguistic sample, we have found a third type of [I]/[V] association, exemplified by German and Dutch and more controversially by Italian and other Free Inversion languages. I have argued that in such languages lexically derived [I,V] constituents are inserted in VP.

This gives us two types of lexical association (which we have called "I-strong" and "V-strong"), exemplified by English modals and tensed verbs in German, respectively; a syntactic V-raising rule, exemplified in English, French, and Scandinavian languages; and a PF lowering rule, exemplified in the same languages. This distribution is summarized in (118):

(118)

	Lexical	Syntactic	PF
Romance	V-strong	-	Lowering
French	-	Raising	-
English	I-strong	Raising	Lowering
Scandinavian	-	Raising	Lowering
Germanic	V-strong	-	-

Three questions immediately arise with respect to (118).

First, why should PF-attachment involve <u>lowering</u> rather than raising ? We have an explanation for why syntactic attachment should involve raising, based on the ECP; but none for the lowering rule wehich seems to characterize PF-attachment. At

the moment, I am not in a position to answer this question; possibly, there is a PF-principle which prefers movement of an affix to a head than vice-versa, but I will not pursue this possibility here, simply leaving the question open for future reaerch.

Second, are the possibilities represented in (118) above independent of each other, or do they cluster in significant ways ? One prediction that we can make with confidence is that languages will either be I-strong (like English); V-strong (like German); or neither (in the case of languages such as French and Norwegian, which do not have lexical [I]/[V] association); but not both. This puts German and English at opposite ends of the spectrum with respect to [I]/[V] association. On the other hand, syntactic V-raising and PF I-lowering appear to be relatively independent of each other; a language may have neither (German), both (Norwegian), or one but not the other (French, Italian).

Third, are the possibilities represented in (117) tied to other properties of the languages in question ? The answer to this question is clearly yes. Thus, it is no coincidence that I-strong lexical association, as in English, is lexically restricted to a small sub-class of verbs, whereas V-strong lexical association, as in German, is available for any verb; since thematic features of the verb are effectively suppressed in I-strong association, only verbs which may assign "secondary" θ-roles (such as modals) or none at all (such as <u>do</u>) are available for association in the lexicon. On the other

hand, in V-strong association, thematic features are not suppressed, and any verb may absorb I-features; as we have seen, verbs do not suppress the Case-features or AGR-features of the I-elements which they absorb, but "adopt" them instead. There is thus an inherent asymmetry between I-strong association and V-strong association, based on the different "feature-adoption" possibilities of [I] (a non-lexical category) and [V] (a lexical category). Note also that I-strong lexical association is itself dependent on the breakdown of a general rule of V-raising, such as applies in French or Norwegian (and such as used to apply in English: see Lightfoot 1979, Roberts 1985 and references therein); it follows that a language without lexically restricted [I]/[V] association (such as all those above except English) cannot have I-strong lexical association.

It remains to be asked why it is that English should differ so radically from its closest relatives; which comes down to the question of why V-raising is lexically restricted in English. A possible answer to this question is provided by Pollock (1987), who suggests that chains created by V-raising, just like any other kind of syntactic chain, can be of two types: A-chains and A'-chains. The former involve relations between θ-heads, and create chains subject to the θ-criterion; in the case of a verbal A-chain, there will always be a θ-criterion violation, as long as the verb is a (primary) θ-role assigner, since by an extension of the "CHAIN condition" of Chomsky (1986a.), repeated below as (119), the unique θ-

assigning position of an A-chain should be at its "foot", which it cannot be if the head is a θ-assigner:

(119) CHAIN condition:

> If C = $(a_1,\ldots,a_n)$ is a maximal CHAIN, then a_n occupies its unique θ-position and a_1 its unique Case-marked position.

This means that the only type of verbal A-chain possible will be one involving no (primary) θ-assigner. On the other hand, verbal A'-chains, created according to Pollock by the need for the tense-operator [±PAST] to govern its scope at LF, should have none of these restrictions. Thus we should get two kinds of V-raising: A-type, restricted to non-θ-assigning verbs, and exemplified by have/be raising in English; and A'-type, unrestricted as to verbal type, and exemplified by French V-raising. Pollock further points out that in infinitivals, French exhibits an optional process of avoir/être raising identical to have/be raising in English:

(120)a. Ne pas embrasser Marie serait dommage.
Not to kiss would be a pity

b. *N'embrasser pas Marie serait dommage.

"It would be a pity not to kiss Marie."

c. Ne pas avoir le courage de l'embrasser serait dommage.
Not to have the courage to OCL her-kiss

d. N'avoir pas le courage de l'embrassser serait dommage.

"It would be a pity not to have the courage to kiss her."

Since no [±PAST] tense operator is present in infinitivals, this is what we should expect. We should also expect have/be raising in English to be optional (except when forced) on the model of the French examples in (120); this prediction seems to be more or less borne out:

(121)a. I seldom have seen such a collection of fools !
b. I have seldom seen such a bunch of twits !
c. I never am happy with the human condition.
d. I am never happy with the price we pay.

Thus French will differ from English in that while the [±PAST] operator must undergo A'-movement in the syntax in French, it will presumably do so at LF in English of course, either possibility is allowed by the principle of Full Interpretation.

Pollock's account thus provides us with a principled approach to the difference between V-raising in French and in English. It should be pointed out, however, that it runs into trouble with Norwegian, since Norwegian has unrestricted V-raising, yet V-raising is not obligatory except when forced by V2. One way round this would be to claim that verbal A'-chains could be formed either in the syntax or at LF in Scandinavian.

A different set of parametric variations characterizes the differences between those languages, such as German and Italian, which allow lexical V-strong lexical [I]/[V] association and those, such as English, French and the

Scandinavian languages, which do not; in 3.61 we suggested that V-strong association might be constrained by a principle of Directional Uniformity, to the effect that Case and θ-roles must be assigned in a single direction by a single constituent. Such a principle predicts that [I,V] may appear in VP in languages with SOV (e.g., German) or VOS (e.g. Spanish or Italian "free inversion" structures), or VSO orders, but not those with SVO orders (e.g., Norwegian, French, English, and SVO structures in the Romance languages), where external Case and θ-roles are assigned in a different direction from internal Case and θ-roles. This prediction appears to be borne out by our admittedly limited language-sample.

The principle of Directional Uniformity thus reduces the possibility of [I,V] constituents in VP to more fundamental (and easily learnable) questions of Case- and θ-role-directionality. Note, however, that while it prevents SVO languages from generating [I,V] constituents in VP, it does not prevent SOV languages, for example, from generating [I] constituents <u>outside</u> VP and employing one of the other [V]/[I] association mechanisms we have examined to satisfy AGR-licensing conditions. Of course, it is an empirical question whether such languages exist; our limited data do not allow us to answer this question one way or the other.

Let me summarize. I have proposed five types of [I]/[V] association, each correlated with a different set of properties, and which take place at different levels of the grammar. (122) presents a summary of these types:

(122)

Level	Type	Restricted	Correlates with
Lexical	V-strong	no	AGR no NP-movement Directional uniformity
Lexical	I-strong	yes	AGR non-θ verbs no A'-type V-raising
Syntactic	A-type	yes	AGR non-θ verbs no A'-type V-raising
Syntactic	A'-type	no	AGR tense
PF		no	AGR

3.73 [I,V] Association and Learnability

Turning more directly to the question of language acquisition, we are now in a position to ask exactly how a child chooses between the various possible options of (122) under the normal circumstances of first-language learning. We will operate from the viewpoint of a child learning English, since this will relate directly to the next section, when we examine data from real-time language acquisition.

A precondition for the setting of any [I]/[V] parameter will obviously be the identification of AGR; in English, this will presumably involve the acquisition of the auxiliary system.

After identifying AGR, the child's next step will be to realize that since English is an SVO language (evidence certainly available at the requisite stage of acquisition),

lexical V-strong association can be dismissed as a possibility, since Directional Uniformity does not hold.

In that case, since [I,V] constituents are not generated in VP, there must be a separate [I] node. The child must now decide whether the language s/he is learning has I-strong lexical association, V-raising, PF-lowering, or all three. Let us assume that PF-lowering is the unmarked case. Since any language with an SAI-type inversion rule (which takes place in the syntax) will have to have either I-strong lexical association or V-raising, as soon as the relevant rule is identified, the child will learn that one or both of them takes place. The same applies to sequences of [NP [I,V] NEG VP].

Next, the child will have to learn whether s/he is dealing with a lexically restricted lexical association/A-type V-raising or an A'-type unrestricted such as that in French. In fact, the logic of learnability imposes a particular structure on such learning, because the language (i.e., the set of strings) generated by the relevant fragment of an English-type grammar is a proper subset of that generated by a Norwegian-type grammar. Thus, whereas <u>any</u> finite verb can occupy the VP-external [I,V] position in French - only a limited subset can occupy such a position in English. This means that were a child learning English to assume as a first hypothesis that English resembled French, only negative evidence, i.e., ungrammatical strings of the form in (123) below, could disabuse him or her of such a hypothesis:

(123)a. *I saw not Bill.

b. *Saw you Bill yesterday ?

In fact, as we shall see in the next section, children learning English never make this type of overgeneralization error, indicating that they never treat English like French. It appears then, that the learner's first hypothesis should be that only a lexically specified subset of verbs may occupy the VP-external [I,V] position. Now, this in turn predicts that children learning French will initially treat French like English - that is, they will use positive evidence to learn which verbs, if any, may raise to INFL. However, this approach also leads to problems, since it seems quite implausible that a French child should have to learn separately for each individual verb the completely general rule of V-raising, which applies to an open lexical class ([V]) rather than a closed functional class (such as the class of aspectual auxiliaries).

In fact, we have already discussed - in 3.5 - the solution to this apparent paradox. There we introduced and justified Emonds' (1985) Designation Convention, which makes a systematic distinction between grammatically relevant features - which characterize closed-class syntactic elements - and purely semantic features, which cannot induce idiosyncratic syntactic behaviour. The Designation Convention is repeated below:

(124) Designation Convention: a rule of syntax (whether of insertion, movement, deletion, or filtering) stated in

> terms of a subcategory of a head of a phrase $[X^0,F]$, where F is a syntactic feature, cannot affect any lexical items under X^0 associated with a purely semantic feature.

Now, the Designation Convention predicts precisely the two types of system exemplified by V-raising in English-type and French-type languages. In the former, the lexically restricted set of verbs (i.e., have and be) which may undergo V-raising may be characterized by syntactic features (and are so characterized in many languages, including Italian (Rizzi 1982, Burzio 1981), Spanish (Torrego 1984) and German (den Besten 1985) where they show "unique syntactic behaviour", as predicted by Emonds. We predict, then, that the child learning English will identify a certain (universally available) set of semantic features as syntactically relevant and switch into a "closed class" learning mode, of the type relevant to inflectional elements; which, as we shall argue in the next section, is precisely what does happen in the acquisition of the English auxiliary system.

The child learning French, on the other hand, will realize as soon as s/he encounters a verb without a syntactically encodable semantic feature in a raised structure (i.e, in either a negated or an inverted context) that V-raising must apply to the entire open-class [V] category; at this point, s/he will immediately shift into a categorial type of learning, whereby a lexical item need merely be classified as a verb in

order to undergo V-raising. We then predict that in the acquisition of the French auxiliary system, there will be a period of "closed-class" learning, followed by a categorial generalization leading to the identification of V-raising as a fully productive syntactic rule.

Finally, the child must decide whether the lexically restricted process in question is a case of lexical association, or restricted V-raising, or both. Once again, learnability considerations dictate that the child's first hypothesis should be the former, since there will be positive evidence, in the form of [Modal[have[be...]]] sequences, for V-raising, but none for the absence of V-raising with modals.

The framework adopted here thus makes rather specific predictions about the course of real-time language acquisition. In the final section of this chapter, we will examine some of these predictions, and argue that while, inevitably, much of the child-language data is interpretable in a number of ways, it is certainly compatible with the approach taken here.

3.8 The Acquisition of the English Auxiliary System

3.80 Introduction

In this section, we will examine the (first language) acquisition of the auxiliary system. Though a considerable literature has grown up around auxiliary acquisition, nearly all of it concerns English, and, as often with child language

literature, much of it is concerned with questions irrelevant to the acquisition of syntax. We will focus on studies which are potentially relevant to our own aims, concentrating in particular on overgeneralizations, which, as we argued in Chapter 1, are important as diagnostics for syntactic "misprojections". In 3.81, we will give an overview of the acquisition of auxiliaries in English; in 3.82, we will focus on a particular set of overgeneralization errors, usually known as "auxiliary overmarking" or "double-tensing" errors. We will not deal in this section with the acquisition of WH-questions or the characteristic class of errors associated with the interaction of SAI and WH-movement, since this will be discussed in detail in Chapter 5.

3.81 Overview of Auxiliary Acquisition in English

In 3.7, we made the prediction that the acquisition of auxiliary verbs in English would precede in a "closed class" manner, due to the child's knowledge that certain types of semantic feature were syntactically encodable, and as such could trigger idiosyncratic syntactic behaviour. In terms of the "closed-class" learning model we are assuming in this thesis (i.e., the Parallel Distributed Processing model of McLelland and Rumelhart 1986), this predicts that the learning of the auxiliary system will be "data-sensitive" and probabilistic; we should therefore expect initial undergeneralization, followed by regularization and

overregularization of "exceptional" cases, followed by an adult system. We should also expect type- and token- variability.

Perhaps not surprisingly, the first part of the prediction is borne out: as observed in naturalistic studies from the Harvard project (Klima and Bellugi 1966, Bellugi 1971, Pinker 1984) to the Bristol study of Language Development (Wells 1979, 1985), and in the experimental studies of Major (1974) and Kuczaj and Maratsos (1975), auxiliary acquisition in English takes place at first in a step-by-step fashion, with little or no generalization from positive to negative auxiliary forms or from non-inverted to inverted contexts. As argued convincingly by Kuczaj and Maratsos, this cannot simply be due to the child's failure to understand that different variants of a given modal are related: both Bellugi (1967) and Kuczaj and Maratsos show, for example, that their subjects understood the relation between contracted and non-contracted forms of the auxiliary will, long before they employed the contracted form productively. Further evidence is provided by the observation of Newport, Gleitman, and Gleitman (1977) that the more parents used yes-no questions when speaking to a child, the faster the child increased the use of auxiliaries in declarative sentences; this clearly indicates that the child understands the connection between the two sentence modalities even though productively s/he does not generalize between them. In other words, even though there is "a great deal of pre-productive integration of the modal system", to quote Kuczaj and Maratsos (1975), there is no overgeneralization, a finding which

indicates children at first employ an inductive learning strategy in acquiring auxiliaries, even when they grasp the generalizations implicit in the system. Again, this is predicted by the model adopted here, where the sets of syntactically relevant and irrelevant semantic features are treated differentially by the first-language learner, the former but not the latter being acquired in a "closed class" fashion.

As reported in the overview given by Stephany (1986), acquisition of auxiliaries proceeds in a relatively invariant order, with the negative modals can't and won't generally appearing before their positive counterparts can and will, which in turn appear before could and would, with should, might, may, and must being acquired last. In fact, Major's study indicates children as old as eight still experience difficulty with the latter group. There are probably several contributing factors to this acquisition sequence, including frequency in the input and semantic complexity; since these are not directly relevant to our concerns, we will not concern ourselves with them here.

The first modal auxiliaries generally make their appearance along with the "semi-modals" hafta, gonna, and wanna; note that though these elements share little in common syntactically with the modals (hafta and gonna are usually analyzed as raising verbs, whereas wanna is a control verb and modals, of course, are neither) they do share semantic features, involving typical "modal" properties of intention, obligation, and volition.

Within the approach taken here, this is readily explained: modal features are amongst those which may be syntactically encodable; if a language does so encode them (as with English modals) then they show unique syntactic behaviour. Now the child learning English knows (as part of UG) that modality features are syntactically encodable; but s/he does not know which of these features are encoded in a given language. It follows that the language learner will employ the same learning strategy (at the same time) for all of them, establishing by positive evidence which verbs appear in which syntactic frames, and ultimately separating the true modals from the semi-modals, which in fact show no syntactic modal behaviour at all.

Thus we see that the generally errorless, "cautious" approach of the first-language learner to the auxiliary system is predicted by the model of the adult grammar we have adopted, and in particular by Emonds' principle of Unique Syntactic Behaviour, itself a consequence of his Designation Convention. Note that a strictly "lexical" theory of learning will also predict such an approach, but cannot differentiate between "closed class" learning of the type discussed here and the productive type of learning which must characterize rules which apply to entire open-class categories.

However, the Parallel Distributed Processing (PDP) learning model we have proposed also predicts a stage of limited overgeneralization to irregular cases. Recalling our discussion in 3.4, we might expect three types of possible error:

(a) "Positional errors", such as those in (125), based on irregular distribution of inverted/uninverted or negated/non-negated auxiliaries:

(125)a. I aren't going to school today !

b. I need go to the toilet !

c. He dare jump over the creek !

(b) "Interpretive errors", based on the absence of a particular reading in an auxiliary position:

(126) May you cut me a piece of bread ?

(c) Inflectional errors, involving over-regularization of irregular variants, such as that in (127), or under-regularization of regular variants, such as those in (128) :

(127)a. I canned do it.

b. He wills go to school.

c. I haved seen it.

(128)a. He don't know.

b. He have two trucks.

c. We was there yesterday.

Given a PDP learning model, the frequency of these errors will be a function of (i) token frequency (the more common an

irregular form, the earlier it will acquired, leading to an "ordering effect" in acquisition, and inducing the typical U-shaped learning curve) (ii) type frequency (the more instantiations of a form there are, the higher the probability that it will influence the decision procedure - thus "regularizations" are more likely than "irregularizations".

In the light of these considerations, let us then examine the three types of possible error identified above.

First of all, let us return to positional errors. Since the overwhelming majority of auxiliary elements (including all the most token-frequent items) have both negative and inverted counterparts, we should expect these errors to be fairly frequent with the few peripheral auxiliaries with positional irregularities. The issue is confused, however, by the fact that at least two of the "positionally irregular" auxiliary verbs, need and dare, have alternate categorial status as regular verbs. In fact, as shown in Major's (1974) study of the later acquisition of the auxiliary system (which we will discuss in more detail in 3.82 below) the overwheming majority of children simply regularize need and dare to the pattern of ordinary verbs taking infinitival complements. As for aren't, the comparative infrequency in the input of examples such as

(129) Aren't I clever !

which are crucial for triggering overegularizations such as (125a.) may explain the infrequency of such errors. However, it

is not at all certain they do not occur; notice also that the relacement form ain't indicates that children are indeed "regularizing" the system.

Turning to "interpetative errors", example (26) and others like it were in fact made by a ten year-old child of my acquaintance, indicating that (a) the predicted error-type occurs, and (b) it occurs late in acquisition, as we would suspect given the sophisticated nature of the rule-inferences involved.

Finally, we turn to by far the most important class of potential errors, those involving the auxiliary inflectional system. Notice that our model makes an initial, very interesting set of predictions: if modals and other auxiliaries are treated by the child as a type of verb, then we should find a "U-shaped learning pattern", with the comparatively token-frequent auxiliaries initially learnt correctly, then overgeneralized as the child learns more regular verbs, leading to errors such as those in (127). On the other hand, if the child treats auxiliaries as a separate class from other verbs, then there will be no overegularization, since the vast majority of the auxiliaries have irregular morphology; instead, we will expect "irregularizations" such as those in (128).

It can be confidently stated that errors such as (127) are absent or almost absent in children's speech, whereas those such as (128) are common. This brings our investigation full circle, since under this interpretation, the development of auxiliaries brings evidence to bear on one of the questions we

asked at the beginning of the chapter - namely, whether a "main verb" or a "modal" analysis is preferable for the adult English auxiliary system. The acquisition data strongly favours the latter hypothesis.

As in Chapter 2, we shall now turn to a "case-study" of a set of errors in acquisition. This time, we will examine a particularly salient and pervasive set of errors, those involving "auxiliary overmarking". It will be argued that the peculiar distribution and nature of these errors can only be understood as the interaction of several independent factors, including the syntactic model of the auxiliary system we have developed in this chapter, the PDP learning model we have adopted, and a particular hypothesis about lexical retrieval.

3.82 Auxiliary Overmarking

First brought to the attention of language acquisition researchers by Hurford (1975), auxiliary overmarking errors have played an important part in the debate between what we might call "grammar-based" and "performance-based" theories of syntactic overgeneralization. In this section we will review the evidence and several hypotheses about its relevance to the developing grammar.

In the appendix to this chapter, I have compiled a list of 119 auxiliary errors, culled from various published sources and several unpublished corpora. While such a list is in some ways

uninformative and even misleading (it does not, for example, tell us anything about individual patterns of error, nor the ratio of errors to grammatical constructions) it does provide us with a data-base from which at least some generalizations can be drawn. It should be emphasized that the present study is as guilty as previous research of the charge of methodological negligence; see Chapter 1 for some suggestions on possible improvements.

In analyzing the data-base for the purposes of auxiliary overmarking, a large number of examples have been excluded from consideration, for one of two reasons. Firstly, as shown by several researchers (including Miller and Ervin 1964, Brown 1973, Kuczaj 1976, and MacWhinney 1982) contracted auxiliaries often remain unanalyzed in the speech of children at the relevant stage of acquisition (for example, in the Harvard corpus, Brown shows convincingly that Adam fails to analyze it's into two morphemes). Thus, in the absence of any way of assessing productivity in the samples from which many of the errors are drawn, I have erred on the side of caution in excluding all contracted auxiliaries. Second, several of the errors in the data-base do not strictly speaking involve "overmarking", since there is only one tense-marking present, though it is realized on the "wrong" tense-bearer; these also have been excluded, restricting the analysis to "double-tensing" errors only. This leaves out an original list of 119 errors, 48 which fall within a strict definition of auxiliary overmarking.

In the typical overmarking error, an auxiliary and a main verb are both marked for tense, leading to (ungrammatical) sentences of the form in (130):

(130)a. What did you bought ?

b. Did you came home ?

c. I did fell when I got blood.

d. It don't hurts.

e. Is it fits ?

f. How can he can look ?

There have been two main types of approach to explaining these errors. The first, exemplified by Hurford (1975), Fay (1978), Mayer, Erreich and Valian (1978), and Klein (1982), uses the characteristics of "intermediate" grammars to account for overmarking; the second, exemplified by Kuczaj (1976), Prideaux (1976), Maratsos and Kuczaj (1978), and Pinker (1984) appeals to other factors, including short-term memory, lexical retrieval, and sentence processing difficulties. We will argue here that all attempts to provide purely grammar-based accounts have failed; but, on the other hand, any attempt to provide a purely grammar-<u>independent</u> account is equally doomed. Instead, it appears that this type of error can only be successfully accounted for by appealing indirectly to grammar, via the mediating effects of performance. This is hardly an earth-shattering conclusion; the reason so few of the other studies have adopted it (once again, Pinker is a notable exception) is

probably connected to the metatheoretical straightjacket worn by those who have taken Chomsky's heuristic distinction between competence and performance as an incontrovertible dogma.

First of all, let us review the grammar-based accounts. Hurford (1975), who initiated the entire debate, is the first to propose that overtensing errors result from the child employing a tense-copying rule rather than a tense-movement rule. Hurford bases his account on Chomsky's original (1957) transformational formulation of SAI, in which a constituent consisting of Tense + {Modal, have, be} is preposed; his claim is that the rule as formulated by the overmarking child simply leaves a copy of the moved constituent in situ. While this original account is clearly inadequate empirically (it fails to account for any overmarking errors which do not involve SAI), its appeal to a decomposition of movement rules into copying and deletion operations has been very influential, culminating in the Basic Operations Hypothesis of Mayer et al. (1978).

Fay (1978) extends Hurford's account. Using Klima's (1964) auxiliary preposing rule for negative declaratives as well as Chomsky's SAI for questions, he claims both are misanalyzed by overemarking children as copying rather than copying and deletion operations.

Mayer, Erreich and Valian (1978) and Erreich, Valian and Winzemer (1980) provide perhaps the most coherent and complete transformational account of overmarking errors, based on what they call the "Basic Operations Hypothesis" (BOH). This hypothesis claims "that, for any transformation which is

composed of more than one basic operation, there exists a class of errors in child speech correctly analyzed as the result of failure to apply one (or more) of the operations specified in the adult formulation of the rule." (Mayer et al. 1978, p.11). It should be noted that the BOH is not only a theory about acquisition, but about adult speech errors, and predicts that auxiliary overmarking should occur (though obviously with less frequency) in adult speech. This prediction is confirmed by Stemberger (1982), who gives examples such as (131) from his corpus of speech errors:

(131)a. What the hell do they do want ?

b. What're we're...What're we gonna get ?

Fay (1980) proposes a similar account of the relation of adult to child speech errors.

The BOH (which subsumes the preceding work of Hurford and Fay) has been subjected to a number of criticisms, both conceptual and empirical. On the conceptual side, contemporary versions of transformational grammar do not analyze movement rules as copying plus deletion operations; in fact, the "basic" operations of copying and deletion are no longer instantiated by any contemporary syntactic rules. On the empirical side, the BOH makes the prediction that other transformational rules will be misanalyzed as copying-without-deletion or deletion-without-copying; while there are some instantiations of these types of errors, there are also notable gaps in the expected error

paradigms, as acknowledged by Erreich et al. (1980), who cite the following types of non-occurring errors:

(132)a. WH-movement : *What did you see what ?
b. ing-hopping : *You could being going to the store.
c. negative-placement : *What not I can't do ?

(Note that (*) in these examples and below indicates an unattested form.)

While the suggestion by Goodluck and Solan (1979) that the BOH does not apply to rules containing "essential variables" will rescue (132a.), it cannot account for the other two types of non-occurring error in (132).

Another problem is created by sentences such as those in (133) which contain two different tenses:

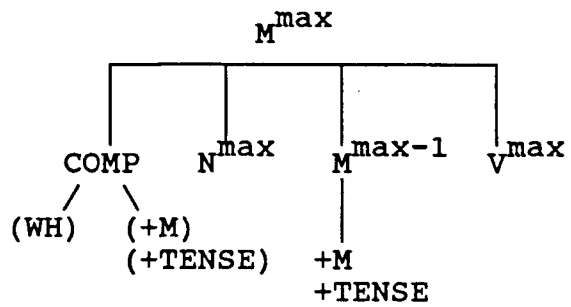
(133)a. What are you did ?
b. Can you broke those ?
c. It didn't has any.

A tense-copying analysis cannot account for such examples. We conclude that the BOH cannot successfully account for auxiliary overmarking.

A rather different grammar-based account of overmarking errors is provided by Klein (1982) who hypothesizes an "intermediate" grammar where auxiliaries are generated both in

COMP and in AUX, as shown in (134):

(134)



(Note: no significance whatsoever is to be attached to the box-like tree-structure in (134), which merely reflects the binary-branching limitations of the word-processor I am writing this on - a good example of an independent performance constraint !)

While such a grammar will generate auxiliary overmarking structures, it raises more questions than it answers. In particular, it predicts that the child will hypothesize a structure like (134) prior to one in which there is only one tense-bearing element. Now, since there is no evidence in the child's input for (134), it will have to be supplied directly by UG as an "unmarked parameter setting". But (134) is highly marked, if not impossible; there are few if any languages which have a "tense-agreement system" whereby every potential tense-bearing element in a finite clause is tensed - in fact, the generalization goes in exactly the opposite way. Without some independent measure of "markedness", Klein's postulated intermediate grammar is simply a description of the data, with no explanatory potential. Moreover, her strictly grammar-based account predicts that auxiliary overmarking should constitute a "stage" in acquisition, which in turn fails to account for why it should occur irregularly in some children's development and

not at all in others', and for why it should also characterize adult speech errors. To be fair to Klein, these last problems are common to all models of syntactic development which postulate such "stages"; we will return to this point below in our discussion of Pinker's (1984) account of overmarking errors. (For more on Klein's analysis, see Chapter 5, where we will discuss her treatment of "non-inversion" errors).

To summarize: no grammar-based account of auxiliary overmarking has achieved either descriptive or explanatory adequacy. This would appear to be a major blow to those of us who on a priori grounds maintain that there must be a connection between syntactic theory and errors in acquisition. However, as I maintained throughout this thesis, such failures do not constitute grounds for abandoning that connection altogether, but rather for abandoning an oversimplified view of the relation between performance-based errors in acquisition and competence-based theories of syntactic structure. This point can be illustrated by the discussion in Mayer et al. (1978) of an earlier hypothesis put forward by Foss and Fay (1975) in which it was claimed that errors such as overmarking were performance-based. Foss and Fay predict that each basic operation performed on an underlying deep structure should incrementally increase the performance difficulty of a given string. This leads to the prediction that overmarking errors should involve fewer operations than grammatical auxiliary structures; but, since in fact overmarking frequently involves do-support and tense-hopping, there is no reduction in the

number of operations performed. Now, Mayer et al. conclude that overmarking errors are competence-based and not performance-based; but in fact, they have shown only that the particular performance model chosen by Foss and Fay is inadequate (not surprisingly, since it assumes a transparent relation between (competence) grammar and its realization in performance in a way strongly reminiscent of the ill-fated Derivational Theory of Complexity (DTC); for arguments that we need not give up the relation altogether simply because of the downfall of the DTC, see Berwick and Weinberg 1984).

Let us then go on to examine other more "performance-oriented" accounts of auxiliary overmarking. First, however, we will introduce a further measure of complexity into the data-base, in the form of certain consistent generalizations about overmarking which no purely grammar-based theory can account for. It turns out that the distribution of overmarking errors is skewed in three ways: first of all, far more overmarking errors occur in non-declarative contexts than in declarative contexts (44 out of 48 in the appended sample); second, more errors occur with complement verbs with irregular tense morphology than with regular morphology (30 out of 48); and third, the vast majority of errors occur with do (41 out of 48). These distributional imbalances have played a significant part in several "performance-based" accounts of overmarking, to which we now turn.

Kuczaj's (1976) response to Hurford's original claim is perhaps the first such account, but its intent is mainly

negative, in that while many arguments are provided against Hurford's analysis, no real alternative is provided. (Kuczaj does, however, point out that cases with contracted auxiliaries are likely to be segmentation errors).

Prideaux (1976) attempts to explain Hurford's data in terms of "functionally-based surface generalizations". The child who overmarks an auxiliary has failed to learn the generalizations repeated in (135):

(135)a. The left-most verb in a clause carries tense.

b. In a clause containing a repeated auxiliary, the first one remains while the second is deleted.

Quite apart from its utter failure to account for the complexities of the adult system (see Chomsky 1975 for arguments relating to the structure-dependence of grammatical rules) Prideaux's analysis is completely uninformative; it merely rephrases the fact that children make errors with auxiliaries. In fact, for Prideaux "there is no need at all... to rely on the untestable and empirically suspect formal notions of deep structures and movement rules" because he fails to understand the difference between a description and an explanation; his analysis makes no predictions about possible errors, nor about the relationship between grammar and performance, nor about how the child uses input information to form the relevant generalizations. Instead, there are vague references to "functional notions", presumably meant to refer to semantic relations between constituents, though at some

points Prideaux seems to be referring instead to grammatical relations such as subject and object. In other words, Prideaux's "analysis" is an example of blinkered empiricism at its worst.

Maratsos and Kuczaj (1978) contains more substantive alternative proposals, though since they are based on an unelucidated theory of lexical retrieval (see Chapter 1), it is difficult to assess them. The basic idea is that there is "a close relation" between the incorrectly chosen form and the correct form, and that "momentary confusion" causes retrieval of the wrong lexical item. Noticing the preponderance of overmarking errors with irregular verbs, Maratsos and Kuczaj conjecture that irregularly inflected (strong) verbs are "more confusable" with their uninflected stems, since both are unaffixed.

While I am sympathetic to this type of analysis, and in fact will adopt a variant of it below, this particular version suffers both from conceptual vagueness and empirical inadequacy. In particular, it predicts that errors of the form in (136) will be as common as those in (137):

(136)a. You couldn't went home.

b. Will he goes ?

c. What wouldn't he did ?

(137)a. I didn't left.

b. What did you got ?

c. Did you had dinner ?

Yet the latter occurs frequently, while the former does not: I have only found one instance, from Eve's eighteenth sample (cited in Pinker 1984, p.265):

(138) You will gone away.

The difference between the two sets of overmarkings is of course that the latter contains do and the former a modal; as predicted above and as we will argue below, this difference falls out from the marked status of do as a modal.

Let us then turn to the most most sophisticated account of auxiliary overmarking, that offered (predictably) by Pinker (1984). The core of Pinker's proposal is that overmarking errors are in fact instances of choosing the "wrong" morphological subcategorization features (constraining equations, in LFG terms) for auxiliaries, which in LFG are simply instances of functional control verbs taking bare stem complements. This makes the prediction that other functional control verbs should also be subject to such errors, a prediction which Pinker claims is borne out by the examples in (139):

(139)a. going being careful

b. Daddy's going taking nap.

c. We've got to pasting.

d. Next year, I'd like to bowling.

e. Make it walks.

f. She gonna fell out.

g. D'you wan' me fell in them ?

h. It's gonna falling near Paul.

i. Lemme sleeps.

j. I can let it spilled.

k. Mommy, d'you wan'me kick you down and broke your crown?

l. Fraser go sitting in that chair.

m. I gonna saw it.

n. Then I'm gonna sit on him and made him broken.

The parallel betweeen such errors and auxiliary overmarking is not nearly as clear as Pinker assumes. First, note that of the fourteen examples, only two are questions and none are negatives; this contrasts very strikingly with the high proportion of non-declaratives in the typical overmarking cases. Second, whereas typical overmarking cases show a high proportion of irregularly inflected verbs, only five out of fourteen of the examples above involve irregular verbal morphology, and six involve the completely regular suffix _-ing_, which is the _first_ verbal inflection acquired by most children (Cazden 1972, Brown 1973); _ing never_ occurs in auxiliary overmarking errors with _do_, and is hardly ever overgeneralized (Kuczaj 1978). Third, errors of the type in (139) are far less common than overmarking with _do_, though, as Pinker points out, more common than overmarking with _be_ or modals. All this casts

considerable doubt on Pinker's claim that auxiliary overmarking is "part of a larger pattern"; rather, it seems that the errors in (139) are part of a separate and much less pervasive pattern.

Let us nevertheless for the sake of argument assume that Pinker is correct in attributing auxiliary overmarking to a failure in complement selection, setting aside the non-parallelism between typical overmarking cases and those in (139) and concentrating henceforth on the former. In that case, an explanation is still needed for why a failure in complement selection takes place. Pinker gives three possible reasons: (a) the morphology of the embedded verb is incorrectly identified; (b) performance factors prevent the constraining equation from being applied; and (c) the equation is not learned to begin with. Pinker's claim is that all three contribute to overmarking; let us accordingly discuss them one by one.

First of all, there is (a), the case of "morphological misanalysis". There are in fact two interpretations of such misanalysis, which we will call "relative" and "absolute". The relative interprewtation closely resembles the "lexical confusion" hypothesis of Maratsos and Kuczaj (1978), which we discussed above; according to this hypothesis, the overmarking child chooses a phonologically, syntactically and semantically closely related form instead of the target form due to some kind of "momentary confusion". It was pointed out above that this failed to account for the preponderance of overmarking errors with do; however, Pinker has an additional explanation

for such errors, based on (c) above; presumably, (a) comes into play more frequently with do since the additional difficulty caused by (c) renders such structures more "prone to confusion".

The absolute interpretation of (a) analyzes morphological misanalysis not as a form of lexical retrieval error but as a failure to learn the correct morphology for an embedded complement verb in the first place. This interpretation is liable to far more serious criticism. In particular, it is subject to the problem we have labelled "token variability". To recall our discussion of Chapter 1, it was pointed out there that different degrees of variability caused different degrees of difficulty for theories of syntactic acquisition, and that token variability, the variable production of the same syntactic structures by the same child, posed particularly acute problems. Now, auxiliary overmarking seems to be characterized by exactly this type of variability; though unfortunately, the anecdotal nature of much of the evidence precludes close examination of the relation of errors to grammatical structures employing the same lexical items, every single study of the phenomenon mentions its sporadic nature. Pinker himself remarks (p.263) that "Many children do not ever make overtensing errors or make them very rarely; no child invariably makes them". Thus it appears likely that the variability of overmarking errors extends to token variability as well as lexical type variability (where individual lexical items show different error patterns). But if so, a mislearning

hypothesis appears extremely unlikely for the morphology of a complement verb: if the morphology had been mislearnt, we should expect a token-consistent pattern of errors.

Thus we conclude that (a) can be a contributing factor to overmarking only under a "relative" rather than an "absolute" interpretation.

Let us then turn to (b), which addresses the role of performance in overmarking errors. Just as with (a), there are two ways of looking at (b): we could either claim that performance factors are responsible for overmarking independently of grammatical structure or we could make the weaker claim that overmarking errors are grammar-dependent types of performance error. Again, let us examine these in turn.

Suppose we were to define "structure-independent" performance effects as those which are based on sentence length and linear position but which make no reference to heirarchical structure (i.e, "sentence-span constraints", as discussed in Chapter 1). It appears most unlikely that such factors, which generally appeal to a simple "length of uttterance" limitation (see 1.32 above) could provide an explanation for overmarking, since there is no correlation between length of utterance and probability of overmarking; most overmarking errors occur in utterances well within the sentence-span capabilities of the children who produce them.

Another type of possible contributing performance factor is that of "interpretive complexity", as discussed in 1.33 above.

Pinker argues that interpetive complexity is indeed relevant to overmarking, citing as evidence the fact (noted above) that overmarking errors occur far more frequently in negative and inverted sentences than they do in declaratives. But in that case, we should expect the sentences in (139) to also contain a high proportion of negatives and questions - quite independently of whether such cases can be considered as parallel in nature to auxiliary overmarking - since, ceteris paribus, we do not expect performance factors to influence one set of structures differently from another. As we have already pointed out, this is not the case.

In fact, any attempt to explain the preponderance of negated and inverted overmarkings with reference to performance factors suffers from several other notable weaknesses. In particular, it is highly unlikely that the negation in elements such as can't, don't, or won't contributes to processing complexity, as predicted by such a view; it is all the more unlikely in lexically-based syntactic theories such as LFG (which, of course, Pinker subscribes to) where positive and negative auxiliaries are treated as separate lexical items.

In fact, there is a far simpler and more satisfactory explanation for the frequency of auxiliary overmarking in negative and declarative contexts. As noted above, the vast majority of overmarking errors occur with do; but do is normally only inserted in inverted or negative contexts (the exception is in emphatic sentences, but these are about as infrequent in discourse contexts as overmarking in declarative

sentences). Thus the preponderance of overmarking errors in non-declarative sentences follows straightforwardly from the distribution of auxiliary do.

We conclude that grammar-independent performance factors cannot be a contributing factor to auxiliary overmarking. This does not, of course, mean that performance is irrelevant; we have taken pains to point out that performance-independent grammatical explanations suffer from equally serious drawbacks, and that the error patterns characteristic of auxiliary overmarking can only be explained as grammatically-based performance effects.

Let us then turn to (c), the hypothesis that the child who makes overmarking errors has failed to learn the constraining equation for the overmarked complement of an auxiliary. Pinker proposes (c) mainly to account for the preponderance of errors with do; since it should be obvious by now that any successful explanation for overmarking must come to terms with this fact, and as Pinker himself points out, no other explanation has been forthcoming, this is a particularly important part of his analysis.

Simply put, Pinker's explanation for the frequency of overmarking errors with do is based on the assumption that the control equation for do is not acquired. His claim is that, since do has no semantic content, the boot-strapping mechanism which enables the child to learn the complement structure of a predicate (Complementation Procedure 4, p.213) fails to apply, and the child misanalyzes do as as "complementless" verb.

Once again, there are two possible versions of this hypothesis, one absolute, one relative. The absolute version claims that the overmarking child has simply failed to learn that do takes any kind of complement, and is subject to the "token-variability" argument which we used against the absolute version of (a): we should expect token-consistency, but once again, the evidence fails to provide it. In particular, the absolute version of (c) makes the prediction that a child who overmarks with do should always do so. But this is not the case; Mayer et al. (1978) point out that their subject R. produced grammatical sentences such as those in (140) at the same time as overmarking errors with do:

(140)a. Why did you say probably ?

b. Do bananas have pits ?

c. Did you cook that ?

We conclude that the absolute version of (c) cannot be maintained.

Let us then turn to the relative version. Here, the overmarking child has learnt that do is a complement taking-verb, but the constraining equation is not "strong" enough to ensure errorless production, and this together with other factors - in particular, the lexical retrieval difficulties discussed in (a) - contributes to the likelihood of auxiliary overmarking.

In support of this hypothesis, Pinker cites one of the experiments of Major (1974); let us accordingly review this

evidence. As reported by Pinker, in a task where children from Kindergarten to Grade 3 were asked to convert declarative sentences with modals into yes/no questions, Major found thirty six instances of children forming questions with do and a modal, of the form in (141):

(141)a. Do you must eat spinach ?

b. Do you ought to fix your lunch ?

c. Do you better go ?

Pinker claims (p.270) that these are "prima facie cases of the child filling both tensed V positions in the matrix clause, one with do, the other with a modal" and that such cases support his contention that because do is "predicateless" (i.e, lacks a semantic argument structure) for overmarking children, it may violate the Uniqueness Condition on f(unctional)-structures (Bresnan and Kaplan 1982) which in LFG prevents two predicates from having the same values in the same sentence.

However, Pinker neglects to report several other significant details of Major's study. First of all, do did not coocur with every modal; on the contrary, the only overgeneralizations of this type that Major recorded were with may (two cases), must (thirteen cases), ought to (thirteen cases), and 'd better (eight cases). Now, these modals are on the periphery of the auxiliary system; it is not at all clear that questions such as those in (142) are even used in colloquial American English:

(142)a. Must you be about six years old ?

b. May you go home with me ?

c. Ought you to fix your lunch ?

d. Had you better go home now ?

Given the rarity of such sentences in the input of the children being studied, it is hardly surprising that they should have difficulty formulating the appropriate questions.

Second, it turns out that while thirty six overgeneralizations of the do + modal type were made, fifty eight were made employing two modals, as in (143):

(143)a. Would you must eat spinach ?

b. Will you might play ?

c. Will you may go ?

d. Would you bewtter go ?

e. Should you ought to fix your lunch ?

Moreover, twenty five errors were of the form in (144), consisting of an aspectual auxiliary plus a modal:

(144)a. Have you must eaten spinach ?

b. Have you better go ?

Rather than providing supporting evidence for Pinker's hypothesis, this is strong counter-evidence, since more than twice as many errors of the relevant type occur with

auxiliaries other than do as occur with do itself.

Thus Major's experimental results, far from supporting Pinker's hypothesis, actually undermine it.

Another counter-argument to the "predicateless do" hypothesis is that it relies crucially on a semantic difference between do and the other auxiliaries. In Pinker's own words, "do is a meaningless verb carrying features for tense and sentence modality." (p.269). But consider an auxiliary such as will, whose sole function is to mark future tense; Pinker's description fits the semantics of will in exactly the same way as that of do. The fact is, the difference between do and the other auxiliaries is not semantic, but syntactic, a fact that his theory is unable to accomodate; we will return to this point below.

We conclude that the hypothesis of "predicateless do" must be rejected, on both conceptual and empirical grounds.

At this point, let us pause and take stock. The following points have emerged from the discussion so far:

(a) Auxiliary overmarking is "performance-mediated"; rather than being an "absolute" effect, it is manifested through a relative increase in speech errors of a type also found in adult speech.

(b) Auxiliary overmarking cannot, however, be simply attributed to grammar-independent performance factors; such factors do not account for the particular characteristics of overmarking

errors.

(c) Auxiliary overmarking is more likely with irregular verb forms. This is plausibly related to difficulties with lexical retrieval.

(d) Auxiliary overmarking is typically associated with auxiliary do. This fact cannot be explained solely by reference to the semantics of do, however.

(e) Auxiliary overmarking is not a subcase of a more general pattern of errors with complement-taking verbs; it shows quite different properties from such errors, and is far more widespread.

(f) Auxiliary overmarking occurs more often in non-declarative contexts; this is almost certainly related to the distribution of auxiliary do, which accounts for most overmarking errors, as noted in (d).

We are now in a position to provide a realistic account of the error patterns characteristic of auxiliary overmarking. Perhaps not surprisingly, we will argue that several factors contribute to this pattern: these include general properties of syntactic errors in language acquisition; the mechanisms of lexical retrieval; and the peculiar syntactic properties of do. Let us discuss these one at a time.

First of all, as we have seen, auxiliary overmarking falls into a class of errors subject to "token variability". As discussed in Chapter 1, such errors cannot be simply instances of syntactic misprojection, and therefore cannot be adequately described by purely grammar-based theories. Rather, we have appealed to properties of the learning mechanism in accounting for them, arguing that the learning of closed-class items should be mediated by a PDP-type "data-sensitive" mechanism.

Next, let us turn to the particular aspect of performance which mediates overmarking: the lexical retrieval mechanism. So far, we have been deliberately vague about the "momentary confusion" which leads an overmarking child to employ an innapropriately tensed verb-form; at this point, I would like to suggest a more explicit account, based on that of Stemberger (1982a.), which is in turn based on work of McLelland and Rumelhart (1981).

What we wish this account to elucidate are the following as yet unanswered questions:

(i) How do we define "lexical confusion" ?
(ii) Why should confusion be more likely to occur with irregularly inflected verbs ?

In formulating answers to these questions, we will draw on two sets of assumptions: the first concerns structure of the lexicon, and the the second the manner in which children learn inflectional morphology.

With respect to the lexicon, we will assume firstly that all members of a given inflectional paradigm constitute separate lexical entries. This is a relatively uncontroversial view, held by most "word-based" morphological theories, from Aaronoff (1976) to Kiparsky (1982). Second, we will assume Stemberger's "interactive activation" model of lexical retrieval; whenever a lexical feature is activated, all lexical entries with the same feature are also activated (see 1.32). Next, we will assume the morphological learning model of McLelland and Rumelhart (1986), described in Chapter 1. Now, we know from the latter that irregularly inflected verbs are "harder" to learn; in PDP terms, the "association lines" which link the inflected and base forms of irregular verbs do not have the same "weight" as those linking the inflected and base forms of regularly inflected forms. Thus, in lexical retrieval, the "activation threshold" for an irregular base-form will be higher than for a regular base-form; hence, retrieval errors involving choice of the wrong form will be more likely (though still only probabilistically) with regular than with irregular verbs.

The account we have just provided makes a number of predictions. First of all, it predicts that auxiliary overmarking need not necessarily involve "double-tensing", as was assumed by the "tense-copying" hypothesis, for example. This prediction is confirmed by examples of overmarking involving different tenses:

(145)a. Don't he wanted to help somebody ?

b. What are you did ?

c. Can you broke those ?

d. It didn't has any.

e. This don't had a nap.

Secondly, since lexical retrieval errors are more likely with irregular verb forms which have not been "consolidated" by the PDP learning mechanism, we can acount straightforwardly for errors such as those in (146) where an "overregularized" past-tense form co-occurs with a tensed auxiliary:

(146)a. They wouldn't haved a house.

b. She didn't goed.

We have now accounted for all the properties of auxiliary overmarking except for one: the preponderance of errors with do. As we argued in our discussion of Pinker's treatment of do, this property is irreducably syntactic; the question is, what syntactic properties of do differentiate it from the other auxiliaries ?

In fact, we have already provided an answer to this question in our discussion of do in 3.36. There it was proposed that do lacked any syntactic features, and thus at NP-structure would simply not exist. This accounted for its "pre-emption" by have/be raising. Now, because do does not exist at the level of syntactic subcategorization (NP-structure) it cannot impose co-occurrence constraints on its "complement", unlike modals,

have, or be; and it is in exactly these circumstances that we should expect children to experience difficulties with respect to which form of the complement verb to insert. Note that any theory which does not make a principled syntactic distinction between do and the other auxiliaries cannot account for this aspect of auxiliary overmarking; we take this as strong (though indirect) supporting evidence for a theory of auxiliaries such as the one which we have presented in this chapter.

Let us summarize. Our account of auxiliary overmarking has the following characteristics:

(a) It is non-unitary. We have appealed to several subcomponents of the theory, including the syntax (for an account of the preponderance of errors with do), the lexicon (for an account of the higher rate of errors wsith irregular verbs) and to a general theory of acquisition (in our discussion of variability). We have argued that such an approach is to be preferred to a purely grammar-based or purely performance-based account.

(b) It is performance-mediated. In other words, error patterns such as overmarking do not bear any privileged direct relation to syntactic theory, but the same indirect relation exemplified by other types of performance data.

(c) It is not performance-based. While the relation between

overmarking errors in acquisition and the theory of syntax is indirect, such errors cannot be dismissed as "performance effects".

What this shows us is that the study of errors such as overmarking <u>can</u> contribute to syntactic theory by providing a significant source of external evidence: but only in conjunction with a viable theory of performance factors in language acquisition. This is both good news and bad news: the good news is that data from language acquisition is potentially relevant to the study of grammar; the bad news is that a good deal of preliminary work must be done before such data can be rendered usable.

3.83 Conclusion

This concludes the first part our examination of the relation between syntactic theory, cross-linguistic markedness and acquisition data in the study of the auxiliary system. In the next chapter we will move onto an extended examination of SAI-type processes; once again, we will begin with a close examination of such structures in English, move onto a cross-linguistic perspective, and finish with an analysis of certain acquisition errors.

APPENDIX: Auxiliary Errors

Hurford (1975)

1. *What's that is ?
2. *What's this is ?
3. What did you bought ?
4. What did you did ?
5. Did you came home ?
6. What did you got ?
7. What did you found ?

Kuczaj (1976)

8. *Is it's Stan's radio ? (N.E., 2;6)
9. *What's he's wearing on his neck ? (A.K, 2;7)
10. Did we ate ? (N.E., 2;5)
11. Did Chris got...? (N.E., 2;5)
12. Did he broke...? (N.E., 2;5)
13. Did we left...? (N.E., 2;5)
14. Did we bought...? (N.E., 2;5)
15. What did I told ? (I.B., 3;1)
16. Don't he wanted to help somebody ? (H.K., 3;6)
17. What did she bought you ? (H.K., 3;6)
18. Did you bought those ? (V.Q., 3;6)
19. *What's was that ? (M.Z., 2;6)
20. What are you did ? (V.Q., 3;7)
21. Can you broke those ? (A.K., 3;0)

22. *I like listen to who's is on this tape. (M.Z., 2;6)
23. *That's makes a truck. (N.E., 2;5)
24. *He's do take his, take his clean pants off. (M.Z., 2;6)
25. *I'm want some dinner. (M.Z., 2;6)
26. *It's don't have any oil in here. (M.Z., 2;6)
27. *What's happen with Santa Claus ? (M.Z., 2;6)
28. *It's looks like a bus. (M.Z., 2;6)
29. *That's are pretty ladies, aren't they ? (V.Q., 3;7)
30. *It's have two. (V.Q., 3;7)
31. *I think it's have a pile (V.Q., 3;7)
32. *Know what's this ? (A.K., 3;11)
33. *I thought so 'cept they're weren't. (A.K., 3;1)
34. *That's means "Get Hawks". (A.K., 3;4)
35. *That what's the witch says to her brother. (A.K., 3;4)
36. I did fell when I got blood. (M.Z, 2;5)
37. You didn't had some. (I.B., 3;1)
38. Don't stuck me. (V.Q., 3;6)
39. They wouldn't haved a house. V.Q., 3;6)
40. I didn't got... (G.D., 3;11)
41. She didn't goed.. (G.D., 3;11)
42. It didn't has any. (A.B., 5;0)
43. You don't has much money. (A.B., 5;0)
44. The plant didn't cried. (A.K., 2;5)
45. It don't hurts. (A.k., 2;8)
46. I didn't saw that. (A.K., 2;9)
47. This don't had a map. (A.K., 2;8)
48. They didn't spilled. (A.K., 3;0)

49. 'Cept you didn't started it, so I started it. (A.K., 3;1

Klima & Bellugi (1966)

50. *Does lions walk ?

Mayer, Erreich, and Valian (1978)

(R., 2;4)

51. Jenni did left with daddy.
52. I did rode my bike.
53. I did broke it.

54. %Is Georgie wake up ? (N., 1;11)
55. *%I'm feed the birdie meat. (N., 2;0)

Ingram & Morehead (unpublished data)

(G.B., 3;0, MLU 3.78)

56. Is it fits ?
57. *%It's go for a ride.
58. *%I'm put some of these on the truck.
59. %Now this one gots stand up.
60. *%This one's go in.

(H.A., 3;1, MLU 5.45)

61. They don't likes to fly.

(B.K., 2;7, MLU 5.08)

62. %I'd better got it.
63. *%I'm put all these ball in.
64. %We put that away, didn't it ?
65. %I might break him, won't I ?
66. %It wasn't hard for me to open for me, is it ?

(R., 2;11, MLU 4.34)

67. *%He's go to the dentist.
68. Why doesn't we has a marble table there ?
69. Why doesn't this goes off ?

(B.R., 2;8, MLU 3.41)

70. *%He's go walking down.
71. *%I'm go pick up something.
72. *%What's you pick up ?

(S.L., 2;9, MLU 5.26)

73. *%He's hold it like that.

74. *%I think this and this doesn't have a broken...

75. *%I gonna putting this one at nursery school.

(A.L., 2;10, MLU 4.50)

76. %Is the elephant fit in it?

77. %I don't got a hurt.

78. %Do you got a hurt ?

79. Did you had wrote that ?

80. *What's you daddy's name is ?

81. *He's gonna see what the lady's do.

82. %Names doesn't work.

83. %*Now he's doesn't throat doesn't hurt.

84. When did you bought these stockings ?

85. *%It's not the teacher's isn't it ?

86. %Let's see we don't got another one.

Johnson data (unpublished).

(Lindsay, 3;0)

87. %Yep, keeping falling off.

(Anthony, 3;0)

88. I didn't missed it.

89. Is this guy goes ?

90. %Watch, here goes the train goes.

91. *He's a wants a go in bed.

Johnson (UBC talk, 1983)

(Kristen, 2;4 - 2;5)

92. %Are you put this on me ?

93. %Are you get this down ?

94. %Are you help me ?

95. %Are you know Lucy's name is ?

96. %Are you want one ?

97. %Are you got some orange juice ?

98. %Are you sneezed ?

99. %Are you hurt yourself ?

100. %Are this the orange juice came from ?

101. %Are this for putting on ?

102. %Are this a big one to carry ?

103. %Are you don't know Sharon's name is ?

104. %Are this is hot ?

105. %Are this is broke ?

106. %Are this is lion ?

107. %Are you write the chimney ?

108. %Are you wake up ?

Menyuk (1969)

109. Where does the wheel goes ?

110. Is this is the powder ? (3;5)

111. Does he makes it ?

112. *What's this plays ? (3;4)

113. What does this does ? (3;5)

114. How can he can look ? (4;4)

115. *%What's these things ? (6;0)

116. *Is that's a belt ? (4;5)r (1984)

117. Does it rolls ?

118. Does it opens ?

119. Did you turned it ?

Abstracting away from examples with contracted auxiliaries (*) and no double marking (%), there are 46 "true" overmarking errors.

CHAPTER FOUR

THE GRAMMAR OF SAI

4.0 Introduction

In this chapter and in Chapter 5 we will be focussing on "SAI-type" processes. The structure of these two chapters will parallel that of Chapter 3, moving from the (synchronic) grammar of English towards a cross-linguistic perspective and finally (in Chapter 5) to an examination of data from first language acquisition. Our primary concerns will also mirror those of Chapter 3. Is SAI in English part of a wider cross-linguistic pattern ? If so, how can cross-linguistic variation be accomodated within a (plausible) model of language acquisition ? And finally, how does such a model relate to data from (real-time) language acquisition ?

Let us begin, then, with a closer look at SAI in English.

4.1 Observations on Subject-Auxiliary Inversion in English

4.10 Introduction

Subject-auxiliary inversion (SAI) is involved in six constructions in English (see Speas 1984). These are

(i) Matrix question-formation

(ii) Tag question-formation

(iii) So-preposing

(iv) Comparatives

(v) Subjunctives

(vi) Negative inversion

In the next six subsections, we will examine each of these constructions in turn.

4.11 Matrix questions

SAI is perhaps most familiar as a component of English question-formation. It is optional in matrix yes-no questions and obligatory in matrix WH-questions, except where a subject has been questioned, in which case inversion leads to ungrammaticality. This pattern is illustrated in (1):

(1)a. You saw Bill yesterday (?)

b. Did you see Bill yesterday (?)*

c. *Who you saw yesterday (?)

d. Who did you see yesterday (?)*

e. Who saw you yesterday (?)

f. *Who did see you yesterday (?)

WH-phrases do occur in fronted position in matrix clauses without inversion. Such clauses are normally interpreted as exclamatives, though the interrogative how come acts exceptionally in not triggering SAI. The set of exclamative WH-phrases is not entirely coextensive with the set of interrogative WH-phrases, though there is overlap, as can be seen below:

(2)a. How silly you are !
   b. How silly are you ?
   c. What a fool you are !
   d. *What a fool are you ?
   e. *How come you left !
   f. How come you left ?

These non-inverted exclamatives should be compared with the closely related so-preposing construction discussed in 4.13 below, which does trigger inversion.

SAI with questions is generally ungrammatical in subordinate contexts:

(3)a. Bill didn't know whether it was still snowing.
   b. *Bill didn't know (whether) was it still snowing.
   c. Bill didn't ask where you had gone.

d. *Bill didn't ask where had you gone.

It should be pointed out that the root/non-root distinction embodied by (3) has been notoriously difficult to formalize. Neither a purely syntactic treatment such as that of Emonds (1976) nor a purely semantic treatment such as that of Hooper & Thompson (1973) seems adequate to account for the range of grammatical and extra-grammatical factors that contribute to the acceptability of sentences where SAI has applied in embedded contexts; see Green (1976) for a description of some of these pragmatic and stylistic complexities. Moreover, the distinction seems to vary in at least three ways. First of all, there is variation across constructions: for example, as we shall see in 4.16, SAI is much more acceptable in embedded contexts with negative preposing than in embedded questions. Second, there is dialect variation: in a number of British and American dialects (see Labov 1968, Stemberger 1982b.) sentences such as (3d.) are perfectly grammatical. Third, there is variation with respect to WH-word: SAI in sentences containing the WH-complementizer whether, such as (3b.), tends to be ungrammatical, even in those dialects which accept inversion in embedded questions headed by other WH-words. Nevertheless, the asymmetry seems clear in the core cases, and we will continue to maintain that any adequate account of SAI must provide an explanation for it.

4.12 Tag questions.

Tag questions share many of the properties of yes/no questions, including an identical root/non-root asymmetry:

(4)a. Bill went home, didn't he ?

b. *I wondered whether Bill went home, didn't he ?

c. *I knew that Bill went home, didn't he ?

They also show similarities with "VP anaphora" processes such as VP deletion; both involve an "empty " VP and a "stranded" tense; both can also optionally strand auxiliary <u>have</u> and <u>be</u>, as in (5):

(5)a. John could have been sunbathing in the Bahamas, but I couldn't (have(been)).

b. John could have been sunbathing in the Bahamas, couldn't he (have(been)).

They differ in several important ways, however. The tense of a tag must be identical to that of the matrix clause; this is not so with VP deletion:

(6)a. John has been swimming in the Carribean, and I may too, one day.

b. *John has been swimming in the Carribean, mightn't he ?

Tags are incompatible with both types of matrix question, as illustrated by (7a.) and (b.); VP deletion seems acceptable with yes/no but not with WH-questions, as in (7c.) and (d.) Note that tag are not incompatible with SAI itself, since they can occur in negative preposing contexts: see (7e.):

(7)a. *Did John say that he would leave early, (or)* didn't he ?

b. *Where did John say he would go, (or) didn't he ?

c. Did John say he would leave early, or did Bill ?

d. *Where did John say he would go, and (where) did Bill ?

e. Never again did John visit the Bahamas, did he ?

Two other properties distinguish tags. First, they display opposite polarity from the matrix sentences to which they are attached:

(8)a. John didn't really rob a bank, did(n't*) he ?

b. John really did rob a bank, did(n't) he ?

As can be seen from (8b.), this opposition is incomplete: positive tags seem acceptable with positive questions.

Second, the NP in a tag is obligatorily a pronoun and obligatorily construed with the subject of the matrix clause. In this tags contrast sharply with VP-deletions:

(9)a. *John could have been rich, couldn't Bill ?

b. John could have been rich and Bill could have been too.

c. John$_i$ could have cheated his partner, couldn't he$_i$?

d. John$_i$ could have cheated his partner, and he$_{*i}$ could have too.

4.13 Comparatives

Inversion in comparatives is optional (and, in fact, stylistically marked):

(10)a. Garbage collectors contribute more than linguists do.

b. " " " " " do linguists.

Comparative inversion clusters with the VP-anaphoric inversion cases in not allowing an overt VP:

(11)a. Linguists produce more garbage than garbage collectors can collect.

b. *Linguists produce more garbage than can garbage collectors collect.

SAI in comparatives appears to show a variable root/non-root contrast: if the comparative clause is embedded along with the clause which controls it, SAI appears to be acceptable, as in (12b.) below; on the other hand, if the comparative clause

is embedded separately from its controlling clause, SAI is unacceptable, as in (12d.):

(12)a. Garbage collectors can earn more than can linguists.

b. My mother claims garbage collectors can earn more than can linguists.

c. Garbage collectors can earn more than John claims linguists can.

d. *Garbage collectors can earn more than John claims can linguists.

We might account for the ungrammaticality of (12d.) by appealing to a locality requirement between the (optionally) SAI-inducing element than and the inverted auxiliary; in (12d.), these two elements are separated by the clause John claims. This would mean that the root/non-root distinction would not after all be relevant to SAI in comparatives.

4.14 So-preposing.

Pace Speas (1984), we will claim here that there are two quite different processes of so-preposing. The first involves a type of quantification parallel to that in exclamatives and related to that in comparatives. This "exclamative so-preposing" fronts any category along with so, as in (13a.- d.), and takes an obligatory result clause, as in (13e.- f.):

(13)a. So stupid was John that he forgot his own name.

b. Such a fool did John appear that nobody took him seriously.

c. So in love was John that he forgot all his troubles.

d. So quickly did John run that he beat everybody.

e. John is so irritating (that I can't bear to talk to him)!

f. So irritating is John (that I can't bear to talk to him)*!

Exclamative so-preposing shows a subject/non-subject asymmetry, as in (14a. -b.), and patterns with comparatives in showing no strong root/non-root asymmetry (14c. -d.):

(14)a. So many politicians (do*) dupe the people, that I despair of democracy.

b. So many people (do)* politicians dupe, that I despair of humanity.

c. So boring was the lecture that John fell fast asleep.

d. John claimed that so boring was the lecture that he fell fast asleep.

The other type of so-preposing, which we will refer to as "conjunctive so-preposing", is more closely related to VP anaphora processes. As in VP deletion and tag question-formation, there is a "missing" VP and a "stranded" tense. Conjunctive preposing also differs from exclamative preposing in not taking an obligatory result clause, as in (15b.); in being paraphrasable by an as clause, as illustrated by (15c.-

d); and in surfacing as neither in negative environments, as shown in (15e.- f.):

(15)a. I flunked the exam and so did Bill (flunk*).

b. John passed the exam and so did Mary (that I was surprised anyone could fail*).

c. I had to resit the exam, and so did most of my classmates.

d. I had to resit the exam, as did most of my class-mates.

e. *I couldn't face the retake, and so couldn't Bill.

f. I couldn't face the retake, and neither could BIll.

Conjunctive so-preposing resembles comparatives and exclamative so-preposing, however, in lacking a strong root/non-root asymmetry:

(16)a. Bill flunked the exam and so did I.

b. Bill knew he would flunk the exam and so would I.

4.15 Subjunctives

Inversion in subjunctives alternates with subjunctive if (again, as a stylistically marked option):

(17)a. If he had the time, John would reread Proust.

b. Had I the time, I would waste it anyway.

SAI in subjunctives shows a weak root/non-root asymmetry:

(18)a. Were I in charge, things would be different.

b. I know that were I in charge, things would be different.

Since inversion alternates with the complementizer if, it is possible to embed if-clauses as well, leading to structures containing two complementizers:

(19) I know that if I were in charge, things would be different.

4.16 Negative preposing

SAI occurs with a set of "affective" adverbs (Klima 1964) (possibly those with "monotone decreasing" properties: see Ladusaw 1981). SAI with adverbs occurs in both matrix and subordinate contexts, showing little or no root/non-root asymmetry:

(20)a. Never had he seen such a snowstorm.

b. Bill said that never had he heard such blatant lies.

As pointed out by Walli (1984) negative preposing is clause-bound: a preposed adverb can never be associated with a lower

clause. Thus, in contrast to (20a.) below, where the WH-word can have wide or narrow scope, the preposed adverb in (20b.) can only be construed with the matrix clause:

(21)a. When did Bill say Mary would leave ?

b. ?Never has Bill said Mary would leave !

Note that negative preposing is much more acceptable when a quantifier-like NP is present in the clause in which it applies; thus the following sentences increase in acceptability from a proper name, through a definite NP and an indefinite NP to a measure phrase:

(22)a. *??Never had John seen Bill !

b. ??Never had John seen the queen !

c. ?Never had John seen a dinosaur !

d. Never had John seen such a beautiful woman !

In fact, replacing the preposed adverb by an NP gives us an identical gradient of grammaticality, though the overall acceptability of the sentences declines:

(23)a. *Bill had John never seen !

b. *??The queen had John never seen !

c. ??A dinosaur had John never seen !

d. ?Such a beautiful woman had John never seen !

4.17 Summary

At this point, let us summarize the characteristic properties of SAI in English, as illustrated in the preceding sections:

(a) SAI involves the syntactic movement of a tensed verb (in our terms, a zero-level [I,V] constituent).

(b) This constituent moves to a position preceding the subject NP but following a finite set of syntactic elements, which include a preposed WH-word (in WH-questions), a so-phrase (in So-preposing cases), than (in comparatives), a complementizer (in subjunctive and negative preposing environments), and a negative phrase or "affective" adverb (in negative-preposing cases).

(c) SAI is usually in complementary distribution with the complementizer whether/if. Where SAI is acceptable in embedded contexts, it is not in complementary distribution with the complementizer that.

(d) SAI shows a varying root/non-root asymmetry: in most dialects of English, SAI is ungrammatical in non-root environments in questions and tag questions, and more acceptable in comparatives, so-preposing cases, subjunctives, and negative inversion cases.

(e) SAI is blocked in structures involving subject-extraction.

(f) Where grammatical, SAI is obligatory in WH-questions, tag questions, so-preposing cases, subjunctives, and negative-preposing cases, but optional (and in fact marked) in comparatives.

The properties detailed in (a. - f.) constitute the observational basis for any treatment of SAI. In the next section, we will review various previous attempts to account for this array of data within the general framework of GB theory.

4.2 SAI and GB-theory

4.20 Introduction

Within a system such as GB which is committed to explanatory adequacy, individual "rules" are an intermediate stage of explanation. The next stage involves the search for higher-order generalizations which might account for why a set of rules exists. In this section, we will review several previous attempts to identify the grammatical principles responsible for the particular array of data we have just examined.

Which GB principles might account for SAI ? Recall the subdomains of GB theory:

(i) θ-theory

(ii) Case theory

(iii) Binding theory

(iv) X-bar theory

(v) Government theory

(vi) Bounding theory

In fact, at least three of these subcomponents have been implicated in the generation of SAI structures; and we shall suggest a fourth. In the following sections we will briefly review previous GB approaches, before going on in 4.3 to propose our own.

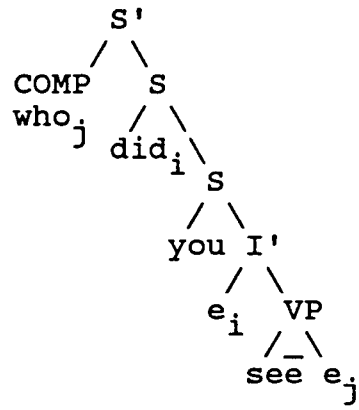
4.21 Safir's HUP

Perhaps the most well-known (and certainly the most extensively criticized) GB account of SAI is provided by Safir & Pestsky (1981) and Safir (1982). The central component of Safir's theory is the Head Uniqueness Principle (HUP) which (roughly) claims that S' must have a unique governed head (INFL is the head of S' in Safir's view). An element in COMP counts as a governor for an adjacent element (where traces, but not PRO, are invisible for the purposes of adjacency). Government is essentially as in the maximally simple Aoun & Sportiche (1983) form; however, a lexical head governs a constituent adjoined to

its maximal projection.

To see how this works, consider a matrix non-subject WH-question:

(24)



The element in COMP counts as a governor for INFL as long as INFL moves into an adjacent position; hence INFL moves, adjoining to S. Since the HUP stipulates that S' have a unique governed head, Safir must further stipulate that the trace of INFL (e_i) is not antecedent governed by INFL itself. This is the core case of SAI. We now need an explanation for the following environments where inversion does not take place, or no element is present in COMP to act as governor:

(a) Subject WH-questions

(b) Non-root contexts

(c) Declaratives

(d) Topicalizations

(e) Exclamatives

(f) Yes/no questions

(a) Subject questions show no inversion, according to Safir, because the WH-phrase in COMP can govern INFL across the trace in subject position (recall that traces do not count for adjacency).

(b) In embedded contexts, INFL is governed by percolation, since it is the head of a governed complement; if INFL were to move to another governed position within an embedded S', it would follow that S would have two governed heads, violating the HUP.

(c) Obviously, INFL is ungoverned in declarative sentences in English. Safir therefore resorts to a universal "Illocutionary Rule"; this rule "looks" for the illocutionary properties of the highest non-empty COMP or INFL, and if the highest node of the relevant type is declarative INFL, exempts it from the HUP (the exact formulation of the I-rule need not concern us here; see Safir 1982, p.438).

(d) The I-rule also exempts topicalization in English from the HUP, since Safir adopts Chomsky's (1977) analysis of topicalization, where the topic does not occupy COMP but a "TOP" node dominated by S"; in that case COMP is "empty" (since it contains only the trace of WH-movement) and the I-rule moves down to INFL, which is declarative, and thus exempted from the HUP.

(e) Exclamatives are relegated to a footnote in Safir (1982); which is to say that he has no explanation for them.
(f) Safir postulates an empty yes/no operator in yes/no questions (essentially equivalent to the abstract Q-morpheme proposed by Baker 1970) which acts as a governor in non-WH questions in a parallel manner to the way in which overt WH-operators govern INFL in WH-questions.

Though Safir's coverage of SAI in English is ingenious and fairly complete, and his theory is extended to inversion rules in French and German, there are a number of technical and conceptual flaws in his approach. Some of these are dealt with in Speas (1984); we will confine ourselves here to the most obvious and serious deficiencies in his analysis.

First of all, note that Safir's approach involves postulating an entirely new sub-component of GB theory, "Inflection Government Theory" (henceforth IGT). Though IGT draws on the central structural relation of GB theory (i.e., government) it cannot be incorporated into any of the other subdomains of the GB framework. This in itself of course does not constitute counterevidence to IGT; but any other theory which <u>can</u> be so incorporated is, ceteris paribus, to be preferred on conceptual grounds.

Second, as discussed in Speas (1984), the IGT notion of government encounters several major technical and conceptual

problems. Safir essentially sets up two distinct notions of government. The first, "percolation government", adapted from Kayne (1984), is a fairly well established form of head-to-head government, which we have adopted (see Chapter 2). The second, which we will call "adjacency government" is a lot more suspicious. To start with, anything in COMP <u>except</u> a lexical complementizer counts as a lexical governor for INFL; this is quite incompatible with the usual view of lexical government, which is a property of X^0 lexical categories. The notion of adjacency used by Safir is also an odd one, firstly because string adjacency plays no role in the usual view of government (it is quite irrelevant to percolation-government, for example), and secondly because it has the peculiar property of ignoring case-marked traces but not <u>PRO</u>; this is precisely the opposite type of adjacency to the one made familiar by the facts of <u>wanna-contraction</u>, where <u>PRO</u> but not Case-marked trace is ignored (see Jaeggli 1980). In fact, Safir is forced to envisage the possibility that different types of adjacency might be relevant to different subdomains of GB theory; while not impossible, this certainly counts as a decrease in the restrictiveness of the theory, and a fortiori an increase in difficulty for the first language-learner.

Thirdly, Safir must assume that a moved INFL does not govern its own trace (or does not leave one at all), since if it did the HUP would be violated. This requires some special

stipulation since it is a requirement of most movement rules that an antecedent govern its trace.

Finally, there is something suspicious about treating ordinary tensed declarative sentences in English as marked with respect to IGT; though the universal I-rule is designed to prevent the language learner from having to find some ad hoc way of circumventing the HUP, it is itself a (universal) ad hoc condition. As we shall argue below, the non-inverted clause order is unmarked, and the inverted order triggered by specific elements in COMP; Safir essentially has the facts the wrong way round.

We thus conclude IGT theory must be rejected as a plausible account of SAI phenomena in English.

4.22 Case-Based Theories of SAI

Let us then turn to another potential source of GB explanation for SAI: the theory of abstract Case. Case theory has proved very profitable as an explanatory tool in several cases of V-movement, particularly those connected with languages which show surface VSO order in tensed clauses but appear to be underlyingly SVO. Welsh (Sproat 1983) is a particularly clear example, since SVO order appears in [-TENSE] complements. The idea is that in such languages (and, in fact in almost all VSO languages), Case-marking is uniformly to the right, but other

principles (possibly connected with θ-role assignment: see Travis 1984) dictate that INFL is generated adjacent to VP. Since INFL is the Case-marker for subjects, it (and, usually, the verb to which it is morphologically attached) moves to an S-initial position in order to assign nominative Case.

Similar explanations have been put foward for languages as diverse as Vata (Koopman 1984), Arabic (Eleseilly 1985), Nishga (Belvin 1984), and Irish (McCloskey 1983). The question is, can such an analysis be extended to SAI in English ?

I think the answer has to be no. INFL in English does not need to move leftward in order to assign Case to its right. In order to extend a Case-marking explanation to inversion in English, the unmarked declarative clause order would have to be specifically exempted from general grammatical principles by some ad hoc condition; the fact is, English is not a VSO language even on the surface. Case-marking INFL-movement shows none of the distinctive properties of SAI; it is not restricted to non-declarative environments, shows no root/non-root asymmetry, and is not triggered by anything in COMP.

A much subtler use of Case-theory in explaining SAI is contained in J. Levin (1983). However, since Case is not crucially implicated in her account, I will defer discussion of it until the next section.

4.23 Operator-Government Theories of SAI

We now move onto what I consider to be the right track in dealing with SAI. Up to now, all the explanations we have been considering have failed to take into account the close relationship between an operator-like element in COMP and INFL-inversion; this relationship is crucial to both of the accounts we will discuss below, as well as our own proposal, which will, however, appeal to a different subcomponent of the system.

First, let us consider the proposals in J. Levin (1983). The basic idea behind Levin's approach is that I-movement is forced by an extended version of the Case Visibility Principle (CVP) which states that

(25) Case is only visible under government.

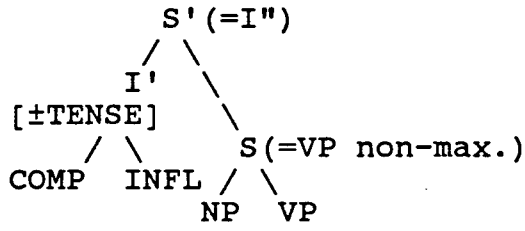
Most versions of the CVP restrict "visibility" to A-chains (Brody 1984); however, Levin extends the notion to A'-chains, which includes operator-variable structures such as those created by WH-movement.

Now, every lexical NP (including WH-phrases in COMP) must have Case to pass the Case-filter, repeated below from Chomsky 1981:

(26) *NP where NP has a phonetic matrix but no case.

(25) and (26) taken together will ensure that every lexical NP must surface in a governed position. Though this has obvious implications for SAI, Levin does not in fact exploit them; this is mainly due to her proposed structure for S', which has the following form:

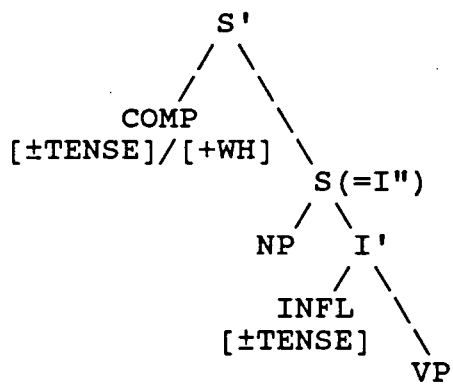
(27)



Since INFL [+TENSE] (=AGR) is a governor, the CVP will be satisfied in any tensed clause, including SAI structures; it cannot then serve to motivate inversion.

Speas (1984) does adapt Levin's insights into an explanation for SAI. Adopting a more orthodox structure for S', she returns INFL to a position within S:

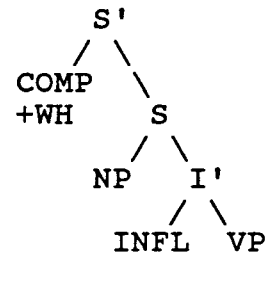
(28)



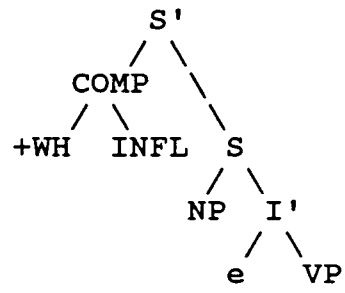
INFL is the head of S; the question remains as to what constitutes the head of S'. Let us first take the case of subordinate clauses. Since Speas makes the assumption that verbs may only select for features of heads of their complements, the answer to this latter question is ambiguous; if the value [+WH] is selected by a verb, then the WH-phrase in COMP is head of S'; if not, then the feature value [±TENSE] acts as "head".

Now, let us turn to matrix clauses. Speas does not discuss the case of declarative clauses; they could either be generated as bare Ss without a COMP, or contain the unmarked feature-specification [+TENSE]. As for clauses where SAI applies, Speas envisages two structures, one prior and one subsequent to INFL movement: these are illustrated in (29a.) and (b) below:

(29)a.



b.



In (29a.), the WH-phrase is the head of S', as in the subordinate clause case discussed above. But what is the head of S' in (29b.) ? In order to avoid a violation of X'-theory,

Speas assumes that INFL "becomes" the head of S' when it moves into COMP.

With these structural preliminaries in place, Speas is in a position to extend Levin's government-based explanation to SAI. She proposes that the following "descriptive principle" holds at S-structure:

(30) THE O-GOVERNMENT PRINCIPLE: Truth-Conditional Operators must be governed.

It is easy to see how this principle accounts for SAI in WH-questions; INFL moves into COMP in order to govern the WH-operator at S-structure; since INFL becomes the head of S' in SAI structures, and since Speas adopts the Aoun/Hornstein/Sportiche version of government, where a head governs everything within its maximal projection, INFL-movement will satisfy O-government with respect to a WH-word in COMP.

Speas goes on to show that <u>all</u> the cases of SAI discussed above involve elements which can be plausibly treated as operators; these include:

(i) Yes/no questions. As in Safir's analysis, an empty Q-operator is assumed, with the form (± x), x a truth value.

(ii) Tag questions. Here Speas makes use of the A-not-A

operator described by Huang (1982) for Chinese. This has the form ((+x) or (-x)), x a truth value.

(iii) So-preposing. The so-phrase is treated as a quantified expression meaning roughly "for some x, x an amount."

(iv) Comparatives. These are said to differ from the so-preposing cases only in that they contain a WH-operator rather than an existential operator: "WHx, x an amount".

(v) Conditional if. Speas shows that conditional if has a negative entailment, unlike non-conditional if, which does not alternate with inversion. She gives the "conditional" operator the following form: not(not x).

(vi) Negative preposing. Negatives have a standard formulation as simple operators: not x.

This is a simplified account; for details, see Speas (op. cit.).

In non-root contexts where SAI does not take place, O-government is satisfied by government-percolation; since a lexical head will govern its complements, it will automatically govern the head of its complements; and since, as we have seen, Speas defines a WH-phrase in COMP as the D-structure head of

S', it follows that a WH-phrase in a subordinate COMP will be O-governed by percolation from a governing lexical head. This explanation also provides a potential answer to the question of why SAI seems better in subordinate clauses with so-preposing, comparatives, subjunctives and negative preposing than it does with questions; the latter but not the former may be subcategorized by particular predicates, which select for features of the heads of their complements.

Thus we see that Speas' account has considerable advantages over the other proposals we have discussed in this section. In particular, it successfully ties SAI to the presence of an operator in COMP (which, as we shall see below in 4.4, is motivated cross-linguistically as well as by English) and provides an explanation for hitherto puzzling differences in root/non-root application of SAI. However, a number of conceptual and empirical problems remain. In the remainder of this section we will discuss these problems, before going on to propose our own explanation for SAI in 4.3

First of all, note that while Speas accounts for the root/non-root distinction, she has nothing to say about the subject/non-subject distinction which characterizes SAI in English. While this is not an inherent flaw in her model, any reasonably complete account of SAI must come to terms with this asymmetry.

Second, while Speas shows that in English it is truth-

conditional operators which trigger SAI, it is certainly not all such operators, as evidenced by the following ungrammatical sentences:

(31)a. *All Chomsky's books have I read.

b. *Perhaps have I told you about this before.

c. *Sometimes have I felt suicidal.

It might be claimed that the operators in (31) are not "truth-conditional"; however, the fact that the French equivalent of perhaps, peut-être, does trigger inversion (see Goldsmith 1981 and section 4.45 below) argues against such a view, unless one wishes to vary the definition of "truth-conditional" from language to language. It appears then that Speas's O-government hypothesis must be weakened to that in (32) below:

(32) O-GOVERNMENT (weak version):

> Certain lexically specified truth-conditional operators must be governed.

However, while empirically more adequate, this version of O-government is still conceptually inadequate in that the notion of "truth-conditional operator" relevant to SAI includes two semantically very different classes of "operator": those with

"distributive" quantifier-like properties (most notably, WH) and those like NEG which are "simple" propositional operators; the question immediately arises as to whether these two types of operator share any semantic property which might allow them to be treated as a natural class for the purposes of SAI.

In fact, there <u>is</u> one intriguing suggestion to this effect, first made by Barwise & Cooper (1981), who make the claim that only downward-entailing/monotone decreasing/antipersistent quantifiers may trigger SAI. The entailments of a downward-entailing quantifer run from set to subset, in contrast to those of an upward-entailing quantifier which run from set to superset; thus in the examples below, from May (1985), the relation between (33a.) and (33b.) is a downward entailment but that between (33a.) and (33c.) an upward entailment:

(33)a. Every man left.
    b. Every father left.
    c. Every man left early.

<u>Every</u> is thus ambiguously upward- or downward-entailing. If we look at negative quantified expressions, we see that they are uniformly downward-entailing:

(34)a. No man left.
    b. No father left.

c. No man left early.

On the other hand, the quantifier many is uniformly upward-entailing:

(35)a. Many men left.

b. Many fathers left.

c. Many men left early.

Now, we are in a position to ask whether all SAI-inducing operators are downward-entailing. In negative preposing cases, the answer is clearly yes; preposed negative adverbials and NPs have the same entailment properties as those in (33). However, when we turn to other SAI-inducing operators, the picture looks much less promising. Consider the following examples:

(36)a. So many men left that the meeting had to be cancelled.

b. So many fathers left that the meeting had to be cancelled.

c. So many men left early that the meeting had to be cancelled.

The relation between (36a.) and (36c.) is unambiguously upward-entailing; that between (36a.) and (36b.) is less clear-cut, but appears also to be upward entailing. In fact, so varies according to the quantifier it is attached to; when it appears

with a downward-entailing operator such as few, then it itself assumes downward-entailment:

(37)a. So few people left that the meeting was packed.

b. So few fathers left that the meeting was packed.

c. So few people left early that the meeting was packed.

Thus we could say that so has no entailment properties on its own; it takes on those of the quantified expression to which it is attached. However, any quantified expression with so can induce inversion:

(38)a. So many people were there at the meeting that I felt like throwing up.

b. So few people were there at the meeting that I went away thoroughly depressed.

We must thus abandon the hypothesis that only downward-entailing operators induce SAI.

Returning to the O-government condition, we conclude that the notion of "truth-conditional operator" as motivated by SAI must include at least the following types of expression:

(i) Most WH-quantifiers, but not exclamatives or how come. This includes comparatives, which are treated as WH-quantified expressions. It might be conceivably extended

to the yes/no question operator, if the latter is treated as a restricted quantification on the domain of truth values (i.e., "for which x, x a truth value").

(ii) Downward-entailing operators; this covers the cases of negative preposing.

(iii) Cases of so-exclamatives, and (more marginally) non-WH exclamatives.

(iv) Cases of conjunctive so, which are presumably treated as logical variants of &.

(v) Conditional if.

It seems highly improbable that the list of expressions above can be characterized in SAI-independent terms as a natural semantic class; but if there is no such characterization, then the notion of "truth-conditional operator" as defined by Speas is merely a paraphrase for "SAI-inducing operator", and has no explanatory value. We are then left with the same question with which we began: why should some operators trigger SAI, but not others ?

One final and important criticism of Speas' proposal is not empirical but conceptual, and concerns the status of O-government within GB theory. Just as with Safir's IGT, it is not at all obvious where O-government fits into the GB picture; must we construct an entirely new O-government subcomponent in order to account for SAI in English ? Such a move seems highly

dubious. The only alternative, however, is to find a way to link Speas' "descriptive generalization" to one or more of the existing components of the system; and this is exactly what we will now attempt to do.

4.3 SAI as A'-type predication

4.30 Introduction

In this section, we will outline our own approach to SAI. Our central claim will be that SAI is related to (syntactic) predication. We will extend the theory of predication developed in Chapter 2 to include not only A-type predication but also what we shall refer to as A'-type predication; this will entail that P-subjects can occupy A'-positions as well as A-positions, but, just as in A-type predication, must be linked to an appropriate AGR-domain; the conditions under which such P(redicate)-linking occurs constitute the core of our theory of predication and, as we shall see, allow us to account for SAI.

In 4.4, we will extend the scope of the analysis to cross-linguistic data, claiming that a theory of predication allows us to formally encode the important pretheoretical distinction between "subject-prominent" and "topic-prominent" languages referred to in much cross-linguistic work, notably the papers collected in Li (1976), and to account for a number of hitherto

puzzling differences between SVO and SOV Germanic languages. In 4.5 we will turn to an explicit parametrization of these differences, which wil prepare us for Chapter 5, in which we will examine relevant language acquisition data, in particular, the well-known "non-inversion" stage in English first-language acquisition.

4.31 A'-type predication

In Chapter 2, we developed a theory of predication based on three important P-licensing conditions and the notions of P(redication)-subject, AGR-linking, and AGR-domain. In this section, we will argue that the elements of A-type predication can be employed to develop a theory of A'-type predication which will provide the basis for an account of SAI-type inversion phenomena. The basic idea is that a lexically restricted set of elements in A'-positions act as A'-type P-subjects in English in non-declarative environments; these are the SAI-inducing "truth-conditional operators" discussed above. It appears (as argued above, and as we shall discuss in 4.33 below) that there is an irreducible element of lexical idiosyncracy involved in determining which fronted operators do and do not act as secondary P-subjects in English; we assume that the lexical entries of truth-conditional operators must include a [±P-subject] feature encoding such idiosyncratic

behaviour.

This is not as conceptually bizarre as it might first appear; in many ways the P-subject-predicate relation is closer to a "topic-comment" relation than to a θ-based relation; see Givon (1976) for arguments that the notion of "subject" is an amalgamam of the discourse function "topic" and the θ-based function "agent". Our claim here will be that languages may define a P-subject in strictly θ-based terms, or in strictly discourse-based (topic-comment) terms, or in both; and that English is of the mixed type. Thus, in declarative clauses in English, predication is satisfied by the external argument of a θ-assigning category, which by definition occupies an A-position (A-type predication); but in matrix interrogatives, predication is satisfied by a fronted WH-phrase in an A'-position, which need not have a primary θ-role and can only receive one through coindexation with an A-position (A'-type predication). It appears to be the case that in "mixed" languages such as English and French certain elements are more amenable to A'-type P-subjecthood than others; in particular, WH-phrases and other SAI-inducing operators may act as P-subjects, while other fronted elements (topics, exclamatives) may not.

Notice that in a language such as English in which predication is predominently θ-based, an A-type P-subject is obligatory even when it receives no θ-role (as has often been

observed; see Chomsky 1982, who incorporates this generalization into the Extended Projection Principle, and Rothstein 1983, who captures it through her version of predicate-linking). On the other hand, A'-type subjects in English are optional, and in fact limited to a small set of syntactic operators. In a language where predication is discourse-based, we might expect the converse to be true; A'-type P-subjects will be obligatory, and A-type P-subjects optional. While we have no evidence for the latter claim, we will show in 4.4 that "V2" languages such as Icelandic and German which have obligatory topicalization and A'-type pleonastic elements (e.g., German es) can be analyzed as instantiations of discourse-based predication in which an A'-type subject is obligatory.

4.32 P-licensing

In Chapter 2, we introduced three P-licensing conditions, repeated below as (39):

(39) I. Every predicate must be contained within an AGR-domain

II. Every AGR-domain must contain a Predication Subject.

III. Every P(redication)-subject must be AGR-linked.

(I) is relevant only to A-type predication; however, (II) and

(III) also apply to A'-type predication. Furthermore, we will argue that languages may differ as to whether (II) applies to both A'-type and A-type predication, or only to A-type predication: this will account for the difference between "discourse-based" and "θ-based" predication alluded to above.

Next, let us turn to the notion of "AGR-linking", repeated below:

(40) AGR-linking: a P-subject α is linked to AGR_i iff α minimally governs AGR_i.

(Recall that minimal government ensures that only the nearest potential governor counts for AGR-linking).

In a (subject-prominent) language with primary A-type predication, linking will surface as syntactic subject-predicate agreement; topic-predicate agreement, on the other hand, is either extremely rare or non-existent even in (topic-prominent) languages with primary A'-type predication, as remarked by Li and Thompson (1976). One plausible explanation for this asymmetry is that A-type predication and A'-type predication are satisfied at different levels of the grammmar: the former is an NP-structure relation, the latter an S-structure relation. If we were to stipulate that morphological agreement rules could only apply at the level of lexical selection (NP-structure) then we could account for the observed

agreement asymmetry between topics and subjects.

Finally, let us turn to the notion of AGR-domain, which we will define in a slightly different way from the version in Chapter 2, incorporating the notion of barrier from Chomsky (1986b.):

(41) α is in the AGR-domain of AGR_i iff α is in the maximal c-command domain of AGR_i, and there is no β, β a barrier for α, such that β excludes AGR_i.

Barrier, itself defined via the notion of Blocking Category defined in (42), is defined as in (43):

(42) Γ is a BC for β iff Γ is not L-marked and Γ dominates β.

(L-marked = θ-governed by a lexical category).

(43) Γ is a barrier for β iff (a) or (b):

(a) Γ immediately dominates δ, δ a BC for β;
(b) Γ is a BC for β, Γ is not equal to IP (= $[I,V]^{max}$).

(For more details and explanation the reader is referred to the discussion in Chapter 2.)

This more sophisticated version of the definition of AGR-

domain will come into play in our discussion of root/non-root asymmetries later on in the chapter.

We are now ready to provide an initial predication-based explanation for SAI. As mentioned above, A'-type predication has a secondary and lexically restricted role in English: certain (SAI-inducing) operators are marked [+P-subject]. When such operators are fronted, they become P-subjects, and must minimally govern AGR, by the P-licensing condition (39.III). The only way to satisfy (39.III) in English is for INFL (containing AGR) to move outside the A-type P-subject; i.e, SAI takes place. To see how this works, take a root non-subject WH-question such as that in (44):

(44) What did you say ?

When the WH-phrase is moved to clause-initial position, a predication relation is immediately established, since fronted WH-phrases in English become A'-type P-subjects. In order to satisfy the predicate-linking rule, INFL must move to a position where it is minimally c-commanded by the subject (see (40)). The only configuration which satisfies these requirements is that created by SAI.

While this is all very well for SAI induced by a WH-phrase or other operator, it does not account for ordinary yes/no questions, which apparently do not have a fronted operator:

(45) Did you take the office key ?

Of couse, we could always claim, following a long tradition stretching back to at least Klima (1964) that yes/no questions do in fact contain a phonologically empty question operator which acts as the trigger for SAI. However, this makes the prediction that yes/no questions are precisely parallel to WH-questions, a prediction which does not appear to be correct, for the following reasons:

(a) To start with, as we shall see in 4.4, there is cross-linguistic evidence that SAI-type inversion in yes/no questions is "parasitic" on SAI in WH-questions; if a language has the former, it will have the latter, but not vice-versa.

(b) SAI in yes/no questions is optional, but obligatory with fronted operators: compare (46b.) with (46d.):

(46)a. Did you go home yesterday ?
b. You went home yesterday ?
c. When did you go home ?
d. *When you went home ?

(c) Yes/no questions have a different intonation pattern from

WH-questions: the rising intonation contour typical of the former is absent in the latter.

In order to account for these differences and at the same time capture the obvious similarities between WH-induced SAI and SAI in yes/no questions, I will adopt and adapt a proposal made by Haider (1986b.). Haider (discussing German V2 structures) suggests that what characterizes "V1" structures in German (which are interpreted as yes/no questions, as in English) is that the fronted "operator-position" normally occupied by a WH-phrase in English, and any XP in German, is filled by the finite verb itself, which acts as a type of yes/no question operator. In our terms, the finite verb (containing AGR) acts as a P-subject; as such it is "vacuously" linked to AGR by (40), satisfying (39.III). An exactly parallel process has been postulated for the "WH-complementizers" <u>if</u> and <u>whether</u>, which satisfy the subcategorization requirements of predicates which take WH-complement clauses while in other ways acting more like [-WH] complementizers (for example in showing *[COMP t] effects and in inducing only "weak" WH-island effects). This analysis effectively captures the ambivalent nature of such elements, without postulating unmotivated empty operators or losing the generalization that SAI is predication-based.

Of course, this is only a bare-bones account of SAI; in

order to achieve even minimal descriptive adequacy, our treatment must at least cover the two main asymmetries which characterize SAI: the subject/non-subject asymmetry and the root/non-root asymmetry.

It turns out that our predication-based theory of SAI accounts for the first of these asymmetries, that between subjects and non-subjects, without any further stipulation at all. Take a root question involving a subject WH-phrase:

(47) Who said that ?

According to our previous assumptions, the WH-phrase in (47) is a P-subject. If so, it must minimally govern AGR. This condition is met, however, without any need for SAI; hence SAI fails to take place. If we adopt the "last resort" theory of movement rules, where movement only takes place when forced by the interaction of grammatical principles, we further account for the ungrammaticality of (48):

(48) *Who did say that ? (non-emphatic reading)

Note that our account of the failure of SAI with "subject extraction" is quite independent of the question of whether the subject in such cases occupies an A- or an A'-position at S-structure. (At LF, it is generally agreed that all WH-phrases

occupy A'-positions). The former hypothesis is argued for in Koopman (1983, 1984); the latter position is adopted by George (1980), Chung and McCloskey (1983) and Chomsky (1986b.). Chung and McCloskey provide evidence that in English no movement takes place; however, we will see in 4.43 below that in several related languages (including Norwegian, Swedish, and Flemish) there is compelling evidence that a subject WH-phrase moves into a pre-sentential position, even in cases of "vacuous movement". This makes the "No vacuous movement" hypothesis first proposed by George and recently adopted by Chomsky highly implausible as a candidate for Universal Grammar. It seems far more likely that the marginal status of A'-type predication in English accounts for the apparent failure of subject WH-phrases to move out of the external argument position (where they can function as A-type P-subjects) and into the pre-sentential position (where they will function as A'-type P-subjects). On the other hand, in "V2" languages such as those mentioned above an A'-type P-subject is obligatory, therefore movement out of an A-position is necessary in order for purposes of predication.

Let us then turn to the second major asymmetry connected with SAI, that between root and non-root clauses. Here we will claim that the relevant distinction is best captured in terms of the notion of AGR-domain introduced in (41). In particular, if we assume that WH-phrases at the head of non-root clauses

are contained in the AGR domain of a higher clause, then we have an explanation for the lack of inversion in such structures, since AGR-domains do not overlap and therefore no element can act as P-subject for an AGR which it c-commands if it is already contained within the AGR-domain of a higher clause.

It will follow that a fronted WH-phrase in a matrix clause will always constitute a P-subject, since it will not be contained in any higher AGR-domain. On the other hand, SAI in embedded clauses will take place just in case the SAI-inducing element is not contained in a higher AGR-domain. This constitutes the core of our analysis of the root/non-root asymmetry; in sections 4.33, 4.43, and 4.44 we will further explore the implications of such an approach.

Thus we see that our predication-based account has the potential to handle the two main asymmetries associated with SAI. We will now turn to a more detailed examination of SAI-related structures in English, focussing on two classes of apparently exceptional elements: those which trigger SAI in unexpected contexts, and those which fail to trigger SAI in expected contexts.

4.33 Exceptional Application of SAI

As we saw in 4.17, only two of the six constructions associated

with SAI in English (questions and tag-questions) show a strong root/non-root asymmetry. The other four (negative preposing, so-preposing, comparatives, and conditionals) do not. We suggested above that the root/non-root asymmetry should be analyzed in terms of the notion of AGR-domain; the question now arises as to whether this distinction adequately differentiates between root and non-root SAI-inducing elements.

There are two parts to our claim: first, we predict that operators which induce SAI only in main clauses are contained in AGR-domains in embedded clauses; second, that operators which induce SAI in both main and embedded clauses are not contained in higher AGR-domains. Let us examine these predictions in turn.

As mentioned above, our explanation for the SAI root/non-root asymmetry is based on the assumption that the [+WH] "head" of an embedded WH-phrase is part of the AGR-domain of the clause in which it is embedded, and thus cannot function as a P-subject. Such an account immediately explains the ungrammaticality of inversion in WH-complements, as shown in (49):

(49)a. I know what you want.

b. *I know what do you want

c. I wondered whether you left.

d. *I wondered whether did you leave.

The ungrammaticality of (49d.) is parallel to that of (49b.) on the assumption that both contain a WH-operator which is a potential P-subject, but which is contained in a higher AGR-domain with its own P-subject.

Note that this explanation cannot easily be extended to relative clauses and other related WH-headed constructions, in which SAI is uniformly ungrammatical, as shown in (50):

(50)a. This is the house which Jack built.
b. *This is the house which did Jack build.
c. It is a house which Jack built.
d. *It is a house which did Jack build.
e. What Jack built was a house.
f. *What did Jack build was a house.

(50a.) and (b.) contain lexically headed relative clauses, (50c.) and (d.) clefts, and (50d.) and (e.) pseudo-clefts. Since by the definition in (41) the WH-phrase in each of these cases is not contained in the AGR-domain of the matrix clause in which it is embedded, it should follow that the embedded WH-phrase can act as a P-subject, triggering inversion. In order to ensure that no inversion takes place in (50) we would be forced to claim that both the NP-head of a relative clause and the fronted WH-phrase with which it is coindexed must be

contained in a higher AGR-domain. This implies that neither the NP head nor the CP which it governs can count as barriers for the purposes of syntactic predication, which in turn means that in terms of Chomsky's (1986b.) definition of barrierhood, CP must count as "L-marked", clearly a counter-intuitive and contradictory proposal.

In order to avoid the problems associated with such a proposal, and in order to account for the cross-linguistic generalization that SAI-type inversion is never triggered by relative-clause type operators, we will assume the following (admittedly descriptive) statement:

(51) A relative-clause type operator can never count as an A'-type P-subject.

We will leave the reasons behind (51) for future research; it is worth pointing out, however, that several researchers have recently proposed that chains headed by relative-clause type operators should be treated in a systematically distinct way from those headed by other WH-operators: see in particular Safir (1986), Contreras (1986), and Tajima (1987).

The notion of AGR-domain thus appears to handle the WH-induced root/non-root SAI asymmetry quite sucessfully. It should be noted, however, that inversion in embedded questions is subject to both dialectal and cross-linguistic variation.

As documented by Labov et al. (1968), from which the following examples are taken, embedded questions in N(on-standard) N(egro) (E)nglish often undergo SAI:

(52)a. You ask him can you play.

b. Let me see could I think of some right away.

c. I don't know how did I do it.

d. You mean how does he gets away.

Moreover, non-root inversion in NNE is not limited to embedded questions, as shown by (53):

(53)a. ...that's what's trumps.

b. If my mother catch me, that's what's the problem gonna be.

It appears, then, that in at least some non-standard dialects (and, as remarked by Labov, probably more often in the casual speech of standard dialect speakers than is generally recognized) SAI does take place in non-root clauses. The question then arises as to how to incorporate such variation into our description of the syntax of SAI, without weakening the analysis itself, a familiar but recalcitrant problem for syntactic theorists concerned with dialectal and sub-dialectal variation.

First of all, note that simply relegating the problem to the

interaction of semantic and pragmatic factors provides no solution. This is because we do not expect dialectal variation in semantic or pragmatic competence; though it is indubitably the case that extra-syntactic factors do influence the likelihood of SAI in embedded contexts (as argued convincingly by Green 1976), such factors are likely to have the same influence on NNE speakers as they do on white middle-class English speakers. This is not to say that syntactic differences between dialects might not allow semantic and pragmatic considerations to have a greater or lesser effect on non-root SAI; rather, such factors cannot themselves determine dialectal variation.

In our model, in fact, dialectal variation of this typehas only one source: language acquisition. Labov's data seem to suggest, then, that the root/non-root distinction is being partially "unlearnt" by children learning English. In 4. ? we will suggest a parameter, based on an dependent/independent clause distinction, which might account for the variable nature of the root/non-root asymmetry.

Let us then turn to cases in which inversion may take place in subordinate clauses. According to the hypothesis developed here, SAI-inducing elements in these cases act as A'-subjects, which in turn means that they are not contained in a higher AGR domain.

As we have seen, elements which trigger embedded SAI can

all appear to the right of a lexical complementizer:

(54)a. I knew that so stupid was he that I could never trust him.

b. I knew that never again would I work with someone that stupid.

c. I knew that had he been slightly more intelligent, everything would have come out rosy.

d. I also knew that he was much smarter than was Bill, his replacement.

(54) appears to provide evidence that, unlike WH-phrases, preposed operators which trigger SAI in embedded clauses adjoin to IP. The alternative is to assume that complementizers may "select" for preposed operators, which in turn trigger inversion; however, it is unclear what "selects" means in that case, since selection imposes no restriction on operator-preposing.

We thus conclude in favour of an adjunction analysis for preposed operators such as those in (54).

Given an adjunction analysis, if we further assume that the AGR domain of a matrix INFL fails to extend past the complementizer of an embedded clause, then we immediately have an explanation for the application of SAI; preposed operators are neither in the AGR domain of the matrix INFL nor of the embedded INFL, and thus become A'-type P-subjects, triggering inversion. Note that in none of the cases in (54) is the SAI-

inducing operator lexically selected by a matrix predicate; this is what we should expect if such elements are outside the AGR-domain of the matrix clause.

Thus our account of the root/non-root asymmetry treats cases of "exceptional inversion" as quite unexceptional - clearly a point in its favour, since as exceptions such cases present obvious learnability problems: a language-learner would need degree-1 positive evidence (i.e., evidence only available in subordinate clauses) in order to ascertain whether SAI was applicable in non-root contexts.

4.34 Exceptional non-application of SAI

We now turn to cases where, according to the theory so far, SAI should be expected to take place, but does not. There are two main subcases to consider, topics and exclamatives. Let us deal with them in turn.

As is well-known, topicalization in English does not trigger SAI:

(55)a. John (*do) I know only too well.

b. On Tuesdays (*do) we go to the cheap movies.

This is a language-specific property of English, as we will see in 4.4 below.

We will assume here that topics, like preposed operators, adjoin to IP (= S). Such an analysis is supported by the fact that in embedded clauses topics appear to the right of lexical complementizers:

(56)a. I know that Mary, Bill admires very much.

b. *I know Mary that Bill admires very much.

One technical problem raised by an adjunction analysis is that though topicalization in English does not trigger SAI, it does display many of the important characteristics of WH-movement, as observed by Chomsky (1977). In particular, it is "unbounded", subject to various subjacency-related island constraints, and displays the *[COMP t] effect, as shown below:

(57)a. John, I know that Bill wants Mary to invite to the party.

b. ??John, I know who invited to the party.

c. *John, I know the man who invited to the party.

d. John, I know (*that) loves parties.

(57a.) is a case of "unbounded" movement; (57b.) a "weak" WH-island violation; (57c.) a "strong" CNPC violation;, and (57d.) illustrates the *[COMP t] effect.

This "unbounded" behaviour is in marked contrast to that of SAI-inducing preposed operators; as observed by Walli (1984),

and noted in 4.16, the scope of such operators is limited to the clause to which they are attached. This is illustrated in (58) below:

(58)a. John said that never again would he eat TV dinners.

b. Never again did John say that he would eat TV dinners.

The operators in (58) have different scope: in (a.), never again has embedded scope, in (b.) it has matrix scope. In neither case is there scope-related ambiguity.

We can explain the contrast between such cases and topicalization if we treat propositional operator preposing as a case of syntactic Quantifier Raising (QR), which is in general bounded (see May 1977); compare the ambiguous sentence in (59a.) with the unambiguous (59b.) and (59c.):

(59)a. Someone loves everyone.

b. Someone thinks everyone loves him.

c. Everyone thinks someone loves him.

In (59a.) either of the quantified expressions someone and everyone may take broad scope; however, in (59b.) and (59c.) only the topmost quantifier may have broad scope. Note that this cannot be an ECP effect, since the ECP would not rule out the missing readings in (b.) and (c.) - the trace of a

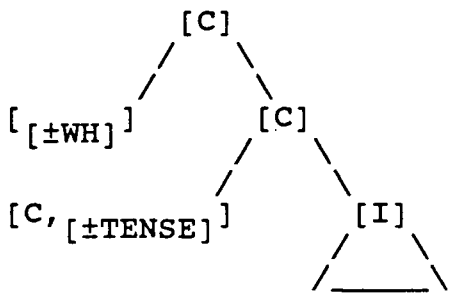
quantifier raised from the subject position of an embedded clause would be antecedent-governed in COMP. Regardless of the true explanation for the boundedness of QR, propositional operator preposing seems to show an identical restriction; we will assume that whatever rules out unbounded QR will also rule out unbounded propositional operator movement. In contrast, unbounded topicalization will be permitted because topicalization is not a case of QR; under standard assumptions, topicalized phrases are "reconstructed" into their original A-positions at LF, and will thus escape the effects of the boundedness restriction.

At S-structure, then, embedded topics will be adjoined to IP, like propositional operators but unlike WH-phrases. Since topicalization is not a case of QR, it is not subject to the QR boundedness constraint.

This is only a partial account of the complex of WH-movement properties illustrated in (57), however; in particular, we have not yet specified the exact mechanisms of "unbounded" movement. There are two main possibilities: successive cyclic adjunction to IP or movement through a peripheral position equivalent to SPEC C, the landing site for WH-movement. It appears that the latter possibility is preferable, in view of the precise parallelism between successive cyclic WH-movement and "unbounded" topicalization; in particular, it seems difficult to account for the *[COMP t]

effect shown in (57d.) without the assumption that there is an intermediate trace in SPEC C which counts as the antecedent-governor for a topicalized phrase extracted from the embedded subject position. This in turn implies that the SPEC C position is available not only for WH-phrases but also for the traces of both topicalization and WH-movement; in terms of the account of phrase-structure presupposed here, the structure of "COMP" will be as in (60):

(60)



A number of questions arise with respect to a structure such as (60). First, we have argued that non-adjoined positions must be licensed at NP-structure by the GREPP; in other words, we cannot simply stipulate the existence of a "SPEC C" position, in the form of a phrase-structure schema, as in Chomsky (1986b.), for example. Second, whereas a WH-phrase may surface in the specifier position, a topicalized phrase never does; there is thus an asymmetry connected with the "lexicalization" of this position.

In reply to the first question, let us first of all suppose

that every clause must be specified for one value of the feature [±WH]. Justification for the [-WH] specification comes from ungrammatical sentences such as those in (61):

(61)a. *I thought what Bill said.

b. *I tried what to do.

We might incorporate this requirement into the grammar as an LF filter; languages without syntactic WH-movement such as Chinese (Huang 1982) require WH-phrases to head [+WH] complements only at LF. Now, there is an asymmetry inherent in the feature-specification [±WH]; whereas the positive value of the feature picks out a set of lexically specified elements, the negative value does not - it is simply the absence of a [+WH] specification. Let us take this absence literally, and claim that the specifier position must be empty in order to bear the value [-WH]. We will then formulate the filter in two parts, as in (62):

(62)a. *[+WH], unless occupied by a WH-operator.

b. *[-WH], if occupied.

Note that condition (62a.) will be met by yes/no questions as well as structures with overt WH-operators, since a fronted finite verb may act as a yes/no operator.

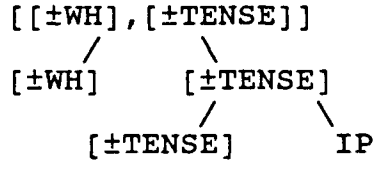
It might be objected that, as we have just shown, intermediate traces of both [+WH] and [-WH] elements must be allowed to occupy the specifier position. However, we have shown only that they occupy this position at S-structure; if we claim, along lines suggested in Lasnik and Saito (1984), that intermediate traces may delete unless required by some principle at a given level of the grammar, and that at the level of selection for [±WH], intermediate traces are no longer required, then the [-WH] position will be empty at LF, as desired.

This means in effect that we are treating *[COMP t] effects as S-structure phenomena; such a view is not consistent with the usual assumption that the ECP (which accounts for *[COMP t] effects) is a condition on LF representation. Note, however, that there are no *[COMP t] effects created by LF-movement (as observed in May 1985, Chapter 5); this has led Contreras (1983) amongst others, to claim that the ECP is best stated at S-structure. We will adopt an intermediate position, claiming that the ECP applies at both S-structure and LF (a position consistent with the hypothesis that there are no conditions which apply only to S-structure), but that it applies to different configurations at each level: *[COMP t] effects are S-structure effects, since (again, following the analysis of Lasnik and Saito) complementizers need not exist at LF. We will also adopt the Lasnik and Saito account of "gamma-marking": in

essence, this involves the assignment of a "proper-government feature" (Γ) to a trace in a properly-governed position, either at S-structure (for A-positions) or at LF (for A'-positions); once an empty category has acquired this feature, it will meet the proper government requirement of the ECP.

Let us further assume that it is the "specifier" position that bears the [±WH] feature at LF. The "COMP" position, on the other hand, is specified for [±TENSE]. This means that every clause will be specified for both features at LF, as in (63):

(63)



(Note that this is an LF-representation; as discussed above in Chapter 2 and elsewhere, there are percolation constraints at NP-structure which prevent tense specifications from percolating if [+WH] is selected as a value for [C]).

So far, then, we have claimed the following: (a) there is a requirement at LF that clauses bear a [±WH] specification; (b) this requirement is met by the pre-complementizer "specifier" position; (c) the [-WH] value is satisfied by an empty specifier; (d) traces may delete unless required at a given level of the grammar; (e) the ECP can be satisfied at either S-structure or LF by the gamma-marking mechanism of Lasnik and

Saito; (f) complementizers may delete at LF.

Next, observe that in English the requirement that a [+WH] element occupy a [+WH] position applies not only at LF (as in Chinese, Japaneses, Malayalam, and other languages without syntactic WH-movement) but also at S-structure. This of course accounts for the ungrammaticality of examples such as those in (64):

(64)a. *I wonder that Bill left who.

b. *I inquired the bus station was where.

Suppose then we were to dissociate the two halves of (62) for languages with overt WH-movement, applying (62a.) at S-structure and (62b.) at LF. This would mean that a [-WH] position could be filled by either a [-WH] or a [+WH] element at S-structure, but not at LF, where it would have to be empty. Now, since intermediate traces may delete unless required, and since they are only required at S-structure, there is nothing to stop such traces from occupying a [-WH] position at S-structure and deleting at LF. On the other hand, lexical elements (either [+WH] or [-WH]) may <u>not</u> occupy the [-WH] specifier position at S-structure, as evidenced by the ungrammatical sentences in (65):

(65)a. *I know Bill that Mary left.

b. *I thought what that she had done.

A plausible explanation for this is that deletion of such elements is non-recoverable (as opposed to the deletion of traces, whose content is recoverable from their antecedents). In that case, a lexical NP in the [-WH] specifier position will be unable to delete subsequent to S-structure and at LF the examples in (65) will be ruled out by the WH-filter in (62b.). Note that we must also rule out reconstruction for sentences such as (65), since otherwise the lexical NP in SPEC C would be able to "return" to its original position, and we would expect no ungrammaticality. We will assume here that reconstruction is restricted to NPs in adjoined positions.

We have thus accounted for the ability of intermediate traces but not [-WH] lexical NPs to occupy the SPEC C position.

To summarize: since only certain lexically specified operators may become A'-type P-subjects in English, topicalized phrases, which adjoin to IP, do not count for the purposes of predication and thus do not trigger SAI. The "WH-like" properties of topicalized phrases are accounted for by postulating a "SPEC C" position which acts as the escape hatch for both WH-phrases and topics; this position must be either occupied by a WH-element or empty at LF, but not at S-structure, where a [-WH] intermediate trace can act as an antecedent governor for the ECP and thus license "unbounded"

movement out of positions which are not lexically governed. "Unbounded" movement of non-WH SAI-inducing operators is prevented by the "boundedness restriction" which appears to constrain QR.

Let us then turn to exclamatives. It might be thought that our account of topicalization could be extended directly to exclamatives. However, it turns out that in many ways exclamatives behave more like WH-phrases than topics, and in some ways like neither. To start with, while no verb ever selects for a topicalized complement, exclamative complements are routinely selected, as argued in convincing detail by Grimshaw (1979), and illustrated below:

(66)a. I'm surprised at what fools politicians are.
b. *I'm surprised at why we don't start an insurrection.
c. I'm surprised that democracy is dead.
d. I wonder why we don't start an insurrection.
e. *I wondered what fools politicians are.
f. *I wonder that democracy is dead.
g. I declare that democracy is dead.
h. *I declare what fools politicians are.
i. *I declare why we don't start an insurrection.

The paradigm in (66) shows that interrogative, exclamative, and declarative complements must be selected separately, except for

one perhaps surprising exception: those predicates such as be amazed at and be surprised at which take exclamative but not interrogative complements also take that clauses, but this is not necessarily true of predicates which take interrogative but not exclamative complements, such as inquire and wonder. Subcategorization for exclamatives also differs from subcategorization for interrogatives in that whereas in the latter case selection of the [+WH] value precludes selection of a value for [±TENSE] (as mentioned above and in Speas 1984), selection of an exclamative complement is only possible when the [+TENSE] value of the [±TENSE] feature is selected. This is illustrated in (67) below:

(67)a. I know what a fool I am.
b. *I know what a fool to be.
c. I don't know what I should do about it.
d. I don't know what to do about it.
e. I know that I should be in therapy.
f. *I know to go into therapy.

The verb know subcategorizes for interrogative, exclamative, and declarative complement clauses. When an interrogative complement is selected, either an infinitival or a tensed clause is possible; however, when an exclamative or a declarative is selected, only a tensed clause is possible.

In other words, for the purposes of subcategorization, exclamatives appear to pattern with declarative and not interrogative complements.

In view of this, we might claim that exclamatives are in fact a special type of declarative, and that semantic selection of the type suggested by Grimshaw (op. cit.) accounts for "exclamative complement selection". In that case, however, we would expect exclamatives and topicalized phrases to behave much alike syntactically. Unfortunately, this is not the case. Unlike topics, exclamatives can never appear in an embedded complement to the right of <u>that</u>:

(68) I realized (*that) what a fool he was !

This suggests that exclamatives, like interrogative WH-phrases, do not adjoin to IP but instead occupy a SPEC C position at S-structure. In fact, exclamatives <u>always</u> occupy this position, and never occur in situ, not even in "echo" interpretation:

(69)a. *You saw what a fool today !

b. *He is how silly !

In this respect exclamatives resemble adjunct interrogative WH-phrases, which are similarly restricted, as shown in (70):

(70)a. *You saw Bill why ?

b. *You fixed the brakes how ?

Thus in some ways exclamatives appear to pattern with other WH-complements, not with declaratives. Note, however, that the set of exclamative WH-phrases is partially disjoint with the set of interrogative WH-phrases, and much more limited, being restricted to what an NP and how AP.

Finally, exclamatives resemble both topics and [+WH] elements in undergoing long-distance movement which obeys the island constraints and shows a *[COMP t] effect:

(71)a. How silly I thought that John realized he had been !

b. ??What a fool Bill wondered who had realized he was !

c. *What a fool Bill knew the man who realized he was !

d. What a monster Bill thought (*that) had arrived to torment him !

It appears, then, that exclamative phrases must be allowed to move through a [±WH] position, like other elements which undergo long-distance extraction.

Thus exclamatives show properties in common with both interrogative WH-phrases and topicalized phrases, but cannot be reduced to either. In fact, if we compare the syntactic behaviour of all the preposed elements we have examined so far, we get the following cross-classification:

(72)	subcategorizable	non-subcategorizable
+SAI	WH-interrogatives	preposed operators
-SAI	WH-exclamatives	topicalized elements

(72) shows that WH-exclamatives occupy the vacant cell in a matrix whose other cells have already been filled. The question is, how do we incorporate their behaviour into the syntactic framework we have been developing ?

To start with, it appears that we must revise our treatment of the syntactic feature [±WH], since exclamatives appear to be [-WH] for the purposes of subcategorization, but [+WH] for the purpose of the LF WH-filter in (62). One possibility is that the feature [±WH] conflates two different types of selection: one involves subcategorization for interrogative or non-interrogative complements (let us call it [±Q]), the other for [±WH] elements, [+WH] in this case including both interrogative and exclamative WH-phrases. Interrogative complements will then be specified [+Q, +WH], ordinary declarative complements [-Q, -WH], and exclamatives [-Q, +WH]. It might be asked what type of element could have the "missing" feature specification [+Q, -WH]; in section 4.46 we will see that a certain set of verbs in Spanish take complements of exactly this type. It is possible that "concealed" questions, as discussed in Baker (1970) and Grimshaw (1979) also fall into this category, though we will not explore this possibility further here.

Turning to the failure of exclamatives to induce SAI, it appears that we must simply stipulate that exclamatives cannot act as P-subjects. As we have already seen, there is an element of lexical idiosyncracy involved in the choice of P-subject in English, which cannot be reduced to either independent syntactic or semantic factors. While it might be argued that exclamatives fall into a natural semantic class, this cannot be said for the deviant WH-phrase how come, which acts syntactically like an exclamative but semantically like an interrogative, and of course fails to induce SAI, as shown in (73):

(73)a. How come you left so early ?

b. *How come did you leave so early ?

However, it is not sufficient to simply claim that P-subjects in English are learnt lexical item by item; such an approach makes incorrect predictions both for the adult grammar (since it cannot account for the subregularities which do appear to characterize SAI-inducing elements) and for language acquisition (since it predicts that the language learner will not overgeneralize SAI, a prediction that is false, as we shall see in Chapter 5). Once again, we are faced with the problem of "near-regularity", a problem which can be resolved neither by adopting a purely "categorical" approach (which cannot account

for exceptions) nor a purely "lexical" approach (which cannot account for regularities).

In Chapter 3, we suggested that the Designation Convention introduced by Emonds (1985) might offer a partial solution to this problem. The Designation Convention, repeated below in (74), makes a principled distinction between closed-class items, which may show "unique syntactic behaviour" and open class items, which may not:

(74) Designation Convention: A rule of syntax (whether of insertion, movement, deletion, or filtering) stated in terms of a subcategory of a head of a phrase $[X^{o},F]$, where F is a syntactic feature, cannot effect any lexical items under X^{o} associated with a purely semantic feature.

(74) predicts that syntactic idiosyncracy is restricted to elements which may be differentiated by purely syntactic features, i.e., closed-class items. This in turn predicts that WH-phrases, which may or may not be P-subjects, are closed-class items. This prediction appears to be accurate: WH-phrases show the other criterial attributes of closed-class elements, showing typically irregular (suppletive) morphological behaviour, and forming a (small) finite set, which may not be augmented by conscious coining.

The Designation Convention thus helps to restrict the

hypothesis space of the language learner; only closed-class items may show unique syntactic behaviour. However, a "lexical" model of language acquisition, even supplemented by the Designation Convention, still cannot account for what appear to be semantic subregularities associated with the application of SAI. As we have seen, interrogative WH-words in general count as P-subjects; exclamative WH-words in general do not. The case of how come shows that this is not exceptionless; but abandoning the prevailing regularity in order to account for exceptional cases does not constitute a viable basis for an explanation. Of course, once again this is a situation where it should be properties of the learning mechanism rather than properties of the grammar which account for the data; as we shall see in Chapter 5, our PDP learning mechanism accounts for both the nature of idiosyncracy in the adult grammar and the observed course of real-time language acquisition.

To summarize: in order to account for the peculiarities of exclamative WH-phrases, we have divided the feature [±WH] into two separate features; [±Q] is the feature selected by predicates which subcategorize either for an interrogative or a non-interrogative complement, while [±WH] is the feature relevant for the LF filter in (62). Exclamatives are specified [-Q, +WH], and [-P], where [±P] represents the ability to act as a subject for the purposes of predication.

4.35 Conclusion

We have now introduced the predication-based theory of SAI and applied it to a wide range of data in English, showing that it accounts straightforwardly for the principle asymmetries associated with SAI-type inversion. Moreover, we have shown that, though there is a residue of idiosyncratic behaviour that cannot be accounted for without stipulation, certain "exceptional" cases of inversion cease to be exceptions under our treatment; this constitutes a clear advantage for our predication-based theory over approaches which are forced to treat such cases as bona fide exceptions.

It is now time to turn to cross-linguistic evidence. Such evidence will provide a further test of the empirical and conceptual adequacy of the predication-based account of SAI, and allow us to identify some of the main parameters associated with SAI-type inversion.

4.4 SAI-type Inversion Cross-linguistically

4.40 Introduction

In this section we will argue that our analysis of SAI in English can be extended to similar constructions in several other languages. These include

(i) "V2" constructions in Germanic languages and elsewhere.
(ii) "Complex inversion" structures in French and other languages with subject clitics.
(iii) "Verb-preposing" structures in Spanish and other Romance languages.

Before examining these cases in more detail, however, we need a working definition for SAI-type inversions. The following criteria, which are neither exclusive nor exhaustive, should serve as a rough diagnostic:

SAI-type inversion rules typically

(a) - involve movement of a tensed [I,V] constituent.
(b) - are triggered by an element in an S-initial position.
(c) - display a subject/non-subject asymmetry.
(d) - display a root/non-root asymmetry.

It is certainly not the case that all SAI-type inversions show all the characteristics above. In fact, if we were to regard SAI as a syndrome, then we could characterize (a.- d.) above as a heterogeneous set of symptoms with varied etiologies. It follows that other "rule-types" might share at least some SAI-type characteristics, and this does indeed appear to be the

case. Take, for example, Case-motivated V-movement, as described in Koopman (1984) and elsewhere. Since this too is a case of V-movement, it shares with SAI characteristic (a.). However, Case-motivated V-movement fails to show any of the other "symptoms" of SAI: it shows no subject/non-subject asymmetry, since it is motivated by Case-assignment, not predication; it shows no root/non-root asymmetry (since Case must be assigned to subjects in subordinate clauses as well as to those in matrix clauses); and it is not triggered by another element in COMP (since V moves to assign Case to an element to its right, not to satisfy predication for an element to its left). Thus while criteria (a.-d.) together serve to delineate SAI-type inversions, none of them alone are necessary or sufficient to characterize SAI.

4.41 General Cross-linguistic Characteristics of SAI-type Inversion

We begin our investigation by repeating Greenberg's (1963) Word-order Universal No.11:

(75) Inversion of statement order so that verb precedes subject occurs only in languages where the question word or phrase is normally initial. This same inversion occurs in yes-no questions only if it also occurs in interrogative

questions.

(75) is intriguing from our point of view for at least two reasons. First, it links SAI-type inversion to an S-external position, which is to be predicted if such rules arise from the need to satisfy predication. Second, it links SAI-type inversion to the presence of an "operator-like" triggering element; this is expected given an analysis such as ours or that of Speas, but quite unaccounted for by the alternative proposed by Safir, for example. Note also that the second part of Universal No.11 suggests that inversion in yes/no questions is "parasitic" on inversion in WH-questions; this is consistent with our treatment of yes/no questions as involving "vacuous" predication, which is contingent on non-vacuous (i.e., operator-induced) predication.

Ultan's (1978) study of question formation in seventy six randomly chosen languages strongly supports Greenberg's claims. Though many of Ultan's assumptions concerning "basic" word order are suspect - he treats Hungarian as an SOV language, for example - both parts of Greenberg's Universal No.11 are supported by his data. Of the thirty eight languages in his survey which have an S-initial WH-word, eight have SAI-type inversion in yes/no questions; of the remaining thirty, none display SAI-type inversion. Ultan notes that in every case where a language has auxiliary verbs, the finite verb is

the one that inverts, giving further support to the idea that it is features of INFL (in particular, AGR) which must appear in inverted structures.

While such statistical surveys as those of Greenberg and Ultan are highly suggestive, they are too vague to provide detailed arguments for or against particular analyses. And so, while noting that our hypothesis is quite compatible with these statistical findings, we move onto a more detailed examination of SAI-type inversions in particular languages, beginning with an examination of the "V2" phenomenon in Germanic.

4.42 "V2 Languages"

First of all, we should point out a misnomer. As we have indicated above, it is not specifically V that occupies second position in V2 languages, but inflected V: therefore this section could more accurately be titled "INFL2".

Second, "INFL2" is by no means restricted to the Germanic languages. Steele et al. (1981) give evidence from Luiseño, a Uto-aztecan language, and Lummi, a Salish language, for an INFL2 condition, instantiated by a particle sequence and a clitic sequence, respectively; Hale (1979) and Nash (1980) report that Warlpiri, an Austronesian language, has only one word order requirement - that the constituent bearing inflection-marking appear in second (or occasionally first)

position in every clause. Of course, without detailed analysis, it cannot be shown that INFL2 in these languages is motivated by the same grammatical principles that motivate English SAI; nevertheless, we find it highly suggestive that languages as diverse as these show an identical restriction on the distribution of sentence-level inflectional elements: this is obviously a topic for further research.

Here, however, we will be restricting ourselves to a detailed look at INFL2 in Germanic, partly due to the accessibility of data, informants, and analyses, partly due to time and space constraints, and partly due to ease of comparison with the (historically related) SAI construction in English.

4.43 INFL2 in German

Germanic languages are characterized by an obligatory [I,V] preposing rule, applying either in main clauses only (as in Dutch) or, variably, in both main and subordinate clauses (as in Icelandic and Yiddish) which places an inflected verb in the second position of a tensed clause, preceded by a maximal projection which is generally analyzed as some form of "topic". Setting aside for the moment those cases where INFL2 occurs in subordinate contexts, let us focus on those languages which show a matrix/subordinate asymmetry with respect to inversion.

In particular, we will concentrate on German, which shows all of the criterial properties of INFL2 languages. In section 4.44, we will extend our analysis to cover parametric variation between the Germanic languages.

Descriptively, we can state the INFL2 condition as in (76):

(76) In all declarative tensed clauses not preceded by a lexical complementizer, a finite verb must occupy second position. Any X" constituent may occupy initial position. In the case of yes/no questions, the finite verb occupies initial position; in WH-questions, the WH-phrase occupies initial position and the finite verb second position.

We illustrate (76) with data from German, beginning with tensed declarative clauses;

(77)a. Gestern hat Hans das Buch dem Herrn gegeben.
Yesterday has Hans the-ACC book the-DAT man given

"Yesterday Hans gave the man the book."

b. Hans hat das Buch dem Herrn gestern gegeben.

c. Das Buch hat Hans dem Herrn gestern gegeben.

d. Dem Herrn hat Hans das Buch gestern gegeben.

All of (77a.-d.) are grammatical, though they differ in emphasis - roughly speaking, the initial constituent is the "topic" of the clause. (78) illustrates a yes/no question:

(78) Hat Hans das Buch dem Herrn gestern gegeben ?
"Did Hans give the book to the man yesterday ?"

(79) illustrates a WH-question:

(79) Wem hat Hans das Buch gestern gegeben ?
"Who did Hans give the book to ?"

Turning to subordinate clauses, the INFL2 condition is inapplicable in German just in case a lexical complementizer is present; as in English, bridge verbs may take a tensed complement without a complementizer, but if they do, INFL2 is automatically triggered, as shown in (80) below:

(80)a. Er sagte, daβ Hans das Buch dem Herrn gegeben hätte.
He said that
"He said that Hans gave the book to the man."

b. Er sagte, Hans hätte das Buch dem Herrn gegeben.

c. *Er sagte, daβ Hans hätte das Buch dem Herrn gegeben.

d. *Er sagte, Hans das Buch dem Herrn gegeben hätte.

Note that the bridge verb sagen takes a subjunctive complement; this constitutes counter-evidence to any analysis which attempts to treat examples such as (80b.) as parentheticals.

This, then, is the INFL2 paradigm. Though, as we shall see,

other Germanic languages differ in some details from the pattern exemplified by German, the basic outline is the same for all of them. We will begin, then, by providing an analysis for the German data, before going onto account for the various differences between the Germanic languages.

First of all, note that INFL2 in German displays most of the criterial properties of an SAI-type inversion: it involves tensed [I,V] movement, is triggered by an S-initial element, and displays a root/non-root asymmetry. (The fourth property, the subject/non-subject asymmetry, which German lacks, will fall out from the obligatory nature of secondary predication in German, as opposed to its optionality in English). In fact, a few minor changes will allow us to extend our analysis of SAI quite straightfowardly to INFL2.

The first of these changes involves what type of phrase can trigger SAI; whereas in English inversion is triggered only by a subset of syntactic operators, in German <u>any</u> preposed phrase can - and must - trigger inversion. In our terms, German is "topic prominent" in the sense that it has obligatory A'-type secondary predication; this is illustrated particularly clearly by the behaviour of the A'-pleonastic element <u>es</u> which is inserted just in case no XP has been fronted, as illustrated in (81) below (from Haider 1986b.):

(81)a. Es irrt der Mensch solange er strebt.

There errs the man as long as he aspires

"So long as Mankind aspires, he errs."

b. Der Mensch irrt solange er strebt.

c. Solange er strebt irrt der Mensch.

Note that pleonastic es cannot be analyzed as occupying an A-type subject position, since, unlike English it or there, it is in complementary distribution with fronted WH-phrases, as shown by (82):

(82) Wer irrt (*es) solange er strebt ?
Who

"Who errs, so long as he aspires ?"

Es is also ungrammatical in yes/no questions, as shown in (83); this follows from our assumption that the fronted [I,V] itself may act as an A'-type P-subject:

(83) Irrt (*es) der Mensch ?

The obligatory nature of secondary predication also accounts for the lack of an SAI-related subject/non-subject asymmetry in German. In English, as we have seen, a WH-phrase in an A-type subject position satisfies predication in situ, since A'-type predication is only optional; in German, on the other hand, A'-type predication is obligatory and thus an A'-type P-subject

must be present in a sentence-external position. Note that German constitutes a clear counter-example to the "Vacuous Movement Hypothesis" of George (1980) and Chomsky (1986b.) since movement of an NP from an A-type subject position to an A'-type subject position (with accompanying [I,V] fronting) is a case of string-vacuous movement.

The second major difference between SAI in English and INFL2 in German concerns the conditions under which inversion is blocked. In (standard) English, as we have seen, SAI-inducing elements fail to trigger inversion just in case they are contained in the AGR-domain of a higher P-subject; WH-phrases, which occupy a "SPEC [P,I]" position, are always contained in a higher AGR-domain, and thus fail to trigger SAI. In this respect, German resembles English, as shown in (84):

(84)a. Ich kann mir denken, was er ihr gesagt hat.
I can refl. imagine, what he to-her said has

"I can imagine what he told her."

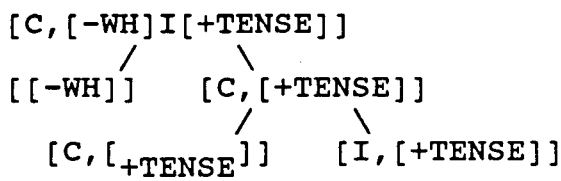
b. *Ich kann mir denken, was hat er ihr gesagt.

However, in embedded declarative clauses, German differs from English in allowing inversion just in case a complementizer is absent (as with bridge verbs; see (80) above). This means that in German, tensed clauses without complementizers function like root clauses, in that they constitute independent AGR-domains which must be supplied with an A'-type P-subject, whereas

clauses with overt complementizers do not. The question then arises as to whether there exists an independent difference between the grammars of English and German which might account for this contrast, or whether AGR-domains must simply be stipulated for individual languages.

In answer to this question, let us first of all review the structure we have proposed for English declarative complements:

(91)



It is usually assumed that where no overt complementizer is present (i.e., optionally in the complements of bridge verbs) a complementizer position still exists. The alternative is to assume that subordinate clauses without complementizers are "bare S" complements, parallel to root clauses. Various theory-internal considerations bear on the choice between these options; in particular, given current theories of government (most notably the "barriers" version proposed in Chomsky 1986b.) the bare S analysis predicts that movement out of a clause without a complementizer should be unconstrained, since S (= IP) can never count as a barrier. However, in fact complementizer-less clauses seem to constrain extraction in much the same way as their equivalents with complementizers

(abstracting away from the *[COMP t] effect, which appears to be connected with lexical properties of the complementizer itself, and is irrelevant to German: see Haider 1986a.). In view of this, it is generally assumed that in English all finite clauses contain a complementizer position, which may be left empty under certain circumstances (possibly when properly governed by a subcategorizing predicate, as suggested in Stowell 1981).

Now, in standard German, there is a strong contrast with respect to extraction properties between finite clauses wuith and without a complementizer. Extraction from a clause containing an overt complementizer is generally ungrammatical:

(86)a. *Wem sagte der Mann, daß die Frau das Buch gegeben
To-who said the man that the woman the book given

hätte ?
has

"Who did the man say that the woman gave the book to ?"

b. *Was sagte der Mann, daß die Frau dem Kind gegeben hätte ?

"What did the man say that the woman gave to the child ?

c. *Wer sagte der Mann, daß das Buch dem Kind gegeben hätte ?

"Who did the man say that gave the book to the child ?"

However, extraction is grammatical out of a clause embedded under a bridge verb without a complementizer which has undergone INFL2:

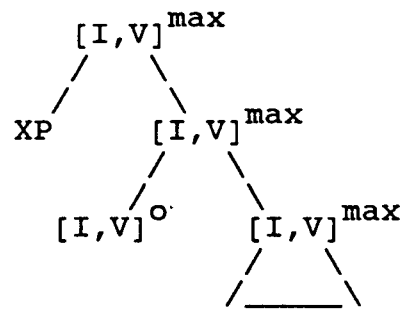
(87)a. Wem sagte der Mann, hätte die Frau das Buch gegeben ?
b. Was sagte der Mann, hätte die Frau dem Kind gegeben ?
c. Wer sagte der Mann, hätte dem Kind das Buch gegeben ?

The examples in (87) appear to contain "INFL1" embedded clauses; however, if we assume that the trace of WH-extraction occupies a preverbal [±WH] position, we can assimilate such cases to the embedded INFL2 pattern illustrated in (80) above. [-WH] phrases may also be extracted, though more marginally:

(88)a. ?Das Buch sagte der Mann, hat die Frau dem Kind gegeben.
b. ?Die Frau sagte der Mann, hat das Buch dem Kind gegeben.
c. ?Dem Kind sagte der Mann. hat die Frau das Buch gegeben.

The contrast between (86) and (87 - 88) shows that there is a significant difference in extraction possibilities between subordinate clauses headed by complementizers and those which have undergone INFL2, a difference which fails to show up in English. This in turn suggests that a "bare S" analysis might be appropriate for embedded INFL2 clauses in German, which would then be structurally isomorphic with root clauses: both would have the structure in (89) below:

(89)



Bridge verbs in German would then take two different sets of finite complement: CP complements, with <u>daß</u>, and [I,V] complements, without. In the latter case, an obligatory A'-type P-subject is required, with attendant [I,V] fronting.

Note that root yes/no questions may <u>not</u> be embedded, even with bridge verbs; the [+WH] complementizer <u>ob/wen</u> is obligatory in indirect questions:

(90)a. Ich fragte, wen die Frau das Buch dem Kind gegeben
I asked whether

hat.

"I asked whether the woman had given the book to the child."

b. *Ich fragte, hat die Frau das Buch dem Kind gegeben.

Since all the bridge verbs which take complementizer-less complements are specified [-Q], this follows without further stipulation.

A number of questions remain, however. The most obvious amongst them is why an [I,V] complement should need an A'-type

P-subject, whereas an [I,V] inside a CP complement apparently does not. Indeed, topicalization with attendant INFL2 inversion inside a CP complement (either declarative or interrogative) is ungrammatical in standard German:

(91)a. Der mann sagte, daβ die Frau das Buch dem Kind gegeben hätte.

b. *Der mann sagte, daβ die Frau hätte das Buch dem Kind gegeben.

c. Der Mann fragte, was die Frau den Kind gegeben hat.

d. *Der mann fragte, was die Frau hat den Kind gegeben.

Once again, we must ask whether our notion of AGR-domain can be extended to cover this contrast.

In answer to this question, let us first of all introduce a distinction between dependent and independent AGR-domains. Obviously, a root sentence will constitute an independent AGR-domain; however, in the unmarked case an embedded complement will not. In a language such as German, which has obligatory A'-type secondary predication, INFL2 will hold in every independent AGR-domain.

First of all, note that the dependent/independent distinction cannot be reduced to a question of lexical selection: relative clauses, in which inversion never takes place, are never subcategorized, yet must be counted as dependent clauses. In fact, as often noted, the

dependent/independent distinction in German seems to be sensitive to the presence or absence of a lexical complementizer (see Goldsmith 1981, Safir 1982, and many others); following Goldsmith, we will refer to this type of distinction as the No Complementizer Condition (NCC). In fact, since WH-phrases which ocupy a pre-complementizer position also block INFL2 - see (90) above - the NCC might better be called the "No COMP condition", where COMP refers to the whole presentential complex, rather than the lexical complementizer itself.

Suppose, then, we were to define the dependent/independent distinction as in (92):

(92) A clause is dependent iff it contains a COMP.

(92) is very likely only one value of a parameter with several other possibilities; we will return to this question later on in this section.

Since the head of a clause is taken to be INFL, we can state (92) as a condition on AGR:

(93) AGR is <u>dependent</u> iff it is contained in an INFL governed by COMP.

(93) allows us to redefine the notion of AGR-domain relevant

for obligatory A'-type predication in a language such as German:

(94) α is in the government-domain of AGR_i iff α is in the maximal c-command domain of AGR_i, where AGR_i is an independent AGR, and there is no β, β a barrier for α, such that β excludes AGR_i.

(99) is a parametric variation of the definition of AGR-domain in (41), incorporating the dependent/independent distinction, which will successfully account for the failure of inversion in clauses headed by complementizers in German.

We have now accounted for the main characteristics of INFL2 in German, using the same predication-based account which we developed for SAI in English. The differences between the two have been reduced to the obligatory and optional nature of A'-type predication in German and English, respectively, and the slightly different definitions of AGR-domain relevant for the two languages.

In the next section we will further explore variation in the INFL2 condition amongst the Germanic languages. It turns out that such variation is quite limited; INFL2 in <u>all</u> Germanic languages (except English) shares the following characteristics:

(a) It involves movement of the finite verb.

(b) It is triggered by any XP in an appropriate S-external position.

(c) It shows no subject/non-subject asymmetry.

(a) follows from the correlation of AGR with finite INFL, and the AGR-linking condition in (40); (b) and (c) are both consequences of the obligatory nature of secondary A'-type predication in all Germanic languages except English.

The main source of variation between Germanic languages with respect to the INFL2 condition involves the "root/non-root asymmetry" (which we have been discussing in terms of the notion of AGR-domain). It is to this variation to which we now turn.

4.44 Variation in INFL2-related Phenomena in Germanic

As mentioned above, there is variation amongst Germanic languages in the applicability of INFL2 topicalization in "non-root" environments. The following table summarizes the distribution of topicalization (with accompanying INFL2 in all languages save English) in such environments in various Germanic languages:

(95)

	[-WH,+C]	[-WH,-C]	[+WH,+C]	[+WH,-C]
German	-	+	-	-
Dutch	-		-	-
Flemish	-		-	-
Frisian	(+)	+	-	-
Yiddish	(+)		-	-
Swedish	(+)*	(+)*	-	-
Norwegian	(+)*	(+)*	-	-
English	(+)*	(+)*	-	-
Icelandic	+		(+)*	(+)*

([+C] refers to a lexical complementizer; thus [-WH,+C] refers to the [-WH] complementizer, equivalent to <u>that</u> in English, and [+WH,+C] refers to the [+WH] complementizer, equivalent to <u>whether</u> in English. ( )* means that the value in question is marked or impossible in some dialects or registers. Note that a blank in the third column indicates that the language in question does not permit the [-WH] complementizer to be absent in non-root clauses).

On the basis of (97), we can distinguish three ways in which INFL2-related topicalization may apply cross-linguistically:

(I) INFL2-related topicalization is restricted to [-WH] clauses without a lexical complementizer (German, Dutch, Flemish, and, as we shall argue below, Frisian underlyingly).

(II) INFL2-related topicalization applies (sometimes marginally, and subject to pragmatic acceptability) in [-WH] non-root clauses with or without a lexical

complementizer (Yiddish, Swedish, Norwegian, English).

(III) INFL2-related topicalization applies (subject to pragmatic acceptability and stylistic considerations) in both [±WH] embedded clauses (Icelandic).

Let us deal with these three cases one at a time, beginning with type I languages which appear, like German, to obey the NCC of Goldsmith (1981).

We have already examined the behaviour of German with respect to INFL2-topicalization. Dutch and Flemish, both SOV languages closely related to German, differ only in that in general lexical complementizers are undeletable, and thus there is no INFL/COMP alternation. Frisian, an SOV language which like German allows a limited range of complementizer-less complements, differs from all three in that in certain embedded complements with lexical complementizers, INFL2-related topicalization <u>is</u> possible; as we shall see, Frisian constitutes an "intermediate" step between the Germanic SOV and SVO languages.

In our discussion in 4.42, we proposed that the notion of AGR-domain relevant for German must incorporate a distinction between dependent and independent AGR, which at least in Type I languages seems itself to depend on the presence of a governing lexical complementizer - see (93) above. However,

while mechanically adequate, this treatment really only restates the NCC in a different theoretical form; a more explanatory approach must come to terms with the question of why it should be that a lexical complementizer should block INFL2-related topicalization in non-root clauses.

As a first step in this direction, suppose we were to claim that in languages which obey the NCC, the complementizer itself is specified for AGR(eement) features. In fact, there is direct support for this view in West Flemish (Bennis and Haegemann 1984) and non-standard dialects of Dutch and German, notably Bavarian (Bayer 1984; see also den Besten 1982), which have complementizers partially inflected for number and gender, as illustrated by the following paradigms:

(96)

	West Flemish	Non-standard German
1sing.	dank	daβ
2sing.	daj	daβste
3sing. (m.)	datje	daβ
3sing. (f.)	dase	daβ
3sing. (n.)	dat	daβ
1plur.	dame	daβ
2plur.	daj	daβste
3plur.	danze	daβ

As can be seen from (96), West Flemish has a far more extensive system of inflected complementizers than non-standard German, where AGR is present only in the second person forms.

However, even where AGR is not generated directly on the complementizer (as in standard Dutch and German dialects) the complementizer serves as a landing site for weak pronoun cliticization, as first demonstrated by den Besten (1977), and illustrated below for standard Dutch:

(97)a. ...dat ie /Piet gisteren die film nog heeft
that he(cl.)/Peter yesterday the film yet has

kunnen zien.
been able to see

"...that he has was still able to see the film yesterday."

b. ...dat gisteren *ie/Piet die film nog heeft kunnen zien.

(97) shows that the "weak" (clitic) pronoun ie is only grammatical when directly adjacent to the complementizer to which it attaches, in contrast to a full NP.

This evidence seems to support the idea that complementizers in the SOV Germanic languages are specified for AGR(eement) features. Now, we have already seen that elements in COMP (including both the complementizer and its [±WH] "specifier") are contained in the government-domain of the AGR of the clause under which they are embedded; this falls out from the definition of government-domain in (41). We can now give a precise characterization of the notion of "dependent" AGR:

(98) AGR is dependent if and only if it is contained in the government domain of another AGR.

In non-SOV Germanic languages (i.e., types II and III above) the complementizer is not specified for AGR; in these languages the AGR features contained in an embedded INFL will not be contained in any higher AGR domain, and will therefore count as independent.

Let us then state the version of (39.III) relevant for Germanic SOV (Type I) A'-type secondary predication:

(99) Every independent AGR-domain must have a unique A'-type P-subject at S-structure.

(99), together with the definition of "dependent AGR" suggested above in (98), accounts for the NCC within the predication-based framework we have been developing.

Before moving on to Type II (SVO) Germanic languages, let us briefly examine "exceptional V2" in Frisian, as discussed in deHaan and Weerman (1986). Frisian in most ways closely resembles other Type I SOV languages: in particular, it exhibits an NCC effect, just like Dutch, Flemish, and German. However, in certain contexts, INFL2-related topicalization in Frisian can take place in embedded contexts <u>with</u> a complementizer, as illustrated in (100):

(100)a. Pyt sei dat hy my sjoen hie.
Pyt said that he me seen had

"Pyt said that he had seen me."

b. Pyt sei dat hy hie my sjoen.

c. Pyt sei hy hie my sjoen.

(100a.) illustrates normal SOV order in a clause headed by the [+TENSE] complementizer dat; (100b.) shows the unexpected variant created by application of INFL2-related topicalization; and (106c.) the NCC effect, where INFL2-related topicalization takes place when the lexical complementizer is absent.

Now, as deHaan and Weerman show, it turns out that the unexpected variant (b) in fact shows many properties in common with the expected NCC variant (c). Firstly, variant (b) is possible in exactly the same contexts as variant (c): that is, with bridge verbs in matrix clauses which are "factual" (i.e., not negated or modalized). This is illustrated in (101):

(101)a. Pyt sei (net) dat hy my sjoen hie.
not

"Pyt didn't say that he had seen me."

b. Pyt sei (net*) dat hy hie my sjoen.

c. Pyt woe sizze dat hy my sjoen hie.
wanted to say

"Pyt wanted to say that he had seen me."

d. *Pyt woe sizze dat hy hie my sjoen.

("Exceptional" inversion also occurs with adverbial degree clauses; since the latter resembles the pattern above in all significant respects, we will focus on the bridge verb cases here: see deHaan and Weerman (op. cit.) for further discussion.)

Second, like Dutch, Frisian has a set of "weak" enclitic pronouns which attach either to a lexical complementizer or a fronted [I,V] constituent. Just as in Dutch, these pronouns may not occupy first position in an INFL2 sentence (presumably, because there is nothing to attach to). Now, it turns out that in variant (b) sentences, the complementizer may <u>not</u> serve as a host for a clitic pronoun; in this respect, (b) sentences resemble (c) sentences, not (a) sentences, as illustrated in (102) below:

(102)a. Pyt sei dat er/hy my sjoen hie.
cl.

b. Pyt sei dat er*/hy hie my sjoen.

c. Pyt sei er*/hy hie my sjoen.

It thus appears that in exceptional inversion contexts, the complementizer <u>dat</u> is "invisible" for the purposes of the NCC. There are a number of ways of capturing this behaviour; deHaan and Weerman opt for a structural solution, in which variant (b) sentences have two complementizer positions. Our own framework,

however, allows us to adopt a rather simpler solution. All we need say is that dat is (optionally) unspecified for AGR features in the complements of verbs which allow exceptional inversion. If the complementizer contains no AGR features, the AGR of the embedded clause will create an independent AGR-domain; by (99) above, such a domain must be supplied with a unique A'-type P-subject at S-structure: therefore, INFL2-related topicalization takes place.

This account also explains the cliticization facts in (102) above, if we make the (natural) assumption that clitic pronouns may only attach to an element specified [+AGR]. We thus see that the properties of "exceptional" inversion in Frisian follow from our approach to INFL2 with minimal parametric modification; Frisian differs from other type I languages only in the option of having finite complementizers unspecified for AGR.

Before leaving Frisian, one other illuminating set of data discussed by deHaan and Weerman deserves brief mention, since it bears directly on the theory adopted here. This is the "en + imperative" construction (which resembles the imperative in form, not function). The essence of their discussion is that clauses headed by the "subordinating conjunction" en contain the specification [-TENSE, +AGR]. Evidence for [-TENSE] is provided by the lack of an independent time reference for the subordinate clause; evidence for [+AGR] by the possibility

for a lexical subject to appear in an en-clause, which deHaan and Weerman assume must be assigned nominative Case by AGR. Now, it turns out that [I,V] movement is obligatory in en-imperatives; this will follow if en, like dat in exceptional inversion cases, is itself unspecified for AGR features, but takes a clause containing independent AGR. As deHaan and Weerman point out, the en+imperative construction provides evidence that the crucial triggering factor in INFL2 structures is AGR, not tense; this is of couse our own position, though many of our theoretical assumptions are very different from those of deHaan and Weerman.

To summarize: Type I (SOV) Germanic languages are distinguished from Type II and III languages by the NCC, which we have derived from the distinction between independent and dependent AGR-domains, plus the proposal that complementizers are specified [+AGR]. These two minor adaptions of the predication-based account of INFL2 have allowed us to account for a number of distinctive properties of Type I languages, lending added support to our theoretical position.

Let us then turn to Type II (SVO) languages, where INFL2-related topicalization takes place in subordinate clauses even when a complementizer is present. Yiddish, Swedish and Norwegian are of this type, though, as noted above, some speakers of Swedish and Norwegian are less than happy with embedded clauses in which INFL2 has taken place.

Type II languages differ from Type I languages in two significant ways. First of all, whereas word-order in Type I languages is SOV in non-INFL2 clauses (i.e., those introduced by a complementizer), in the Scandinavian languages and in Yiddish word-order is normally SVO in subordinate clauses. This is illustrated below with data from Yiddish (Y), Swedish (S), and Norwegian (N):

(103)a. (Y) (Ikh vais, az) er hot geleien dos bukh
b. (S) (Jag vet, att) han har läst boken
c. (N) (Jeg vet, at) hann har lest den bok
(I know that) he has read the book
"(I know that) he has read the book."

There are two ways to account for the word-order differences between Type I and Type II languages. We could either claim that the underlying order of major constituents in the Scandinavian languages and in Yiddish is identical to that in English, i.e S INFL V O, or we could claim that such languages have head initial VPs (i.e VO) but head-final clauses (i.e. S VP INFL) with obligatory INFL2-related topicalization in both main and subordinate clauses. In fact, there is evidence in favour of the former hypothesis from several sources, including adverb placement, the position of negation, and extraction asymmetries.

As pointed out by Platzack (1983, 1986) adverb placement in Swedish shows a "root/non-root" asymmetry (Norwegian shows an identical asymmetry). In root clauses, the finite verb occupies second position, preceding a sentence adverbial, whereas in non-root clauses it is generally found in third position, following an S-adverbial. This is illustrated below:

(104)a. Han (*verkligen) hade verkligen skrivit boken.
He (really) had really written the-book.

"He really had written the book."

b. Jag frågade om han verkligen hade (*verkligen)
I wondered if

skrivit boken.

"I wondered if he had really written the book."

This asymmetry is readily accounted for if we assume S-adverbials are base-generated in a pre-INFL position, and no INFL2-related movement takes place in embedded clauses such as (104b.). A very similar asymmetry characterizes the position of the negative element inte/ikke in the continental Scandinavian languages: in root clauses it occurs after the finite verb, in non-root clauses it precedes the finite verb. This is illustrated below, this time with data from Norwegian:

(105)a. Hann (*ikke) skjønte ikke dette spørsmålet.
He (not) understands not this question.

"He doesn't understand this question."

b. Jeg lurte på om hann ikke skjønte (*ikke) dette
I wondered if

spørsmålet.

"I wondered whether he didn't understand this question."

(105) provides further evidence for an SVO analysis for the Scandinavian languages.

In Yiddish, the same conclusion can be reached on the basis of an extraction asymmetry pointed out in Lowenstamm (1977). It turns out that extraction can only occur out of an embedded clause if the subject is preverbal (examples from Travis 1984):

(106)a. ...der yid velkhn zey hobn geharget in vilne...
the Jew who they have killed in Vilna

"the Jew who they killed in Vilna"

b. *...der yid velkhn in vilne hobn zey geharget...

c. Ikh veys nit vemen zi hot gezen zuntik
I know not who she has seen Sunday

"I don't know who she saw on Sunday."

d. *Ikh veys nit vemen zuntik hot zi gezen.

Note that the ungrammaticality of (106b.) and (d.) is not the result of a general constraint against topicalization in embedded contexts; as we have already mentioned, and as we shall see below, no such constraint exists in Yiddish. The asymmetry in (106) is most easily accounted for if we assume

that embedded topicalization creates a syntactic island (precisely the same claim is made for Frisian by deHaan and Weerman 1986); if so, the examples in (106a.) and (c.) cannot involve topicalization, which leads us back to the conclusion that Yiddish has the underlying constituent order S INFL VP.

Having then established SVO order for Type II languages, let us return to the question of INFL-2 related topicalization in non-root contexts. As mentioned above, the process is somewhat marginal in the continental Scandinavian languages, as illustrated below in (107):

(107)a. (N) Vi vet (at) Jon ikke kysset Lisa.

b. (S) Vi vet (att) Jan inte kysst Lisa.

We know (that) John not kissed Lisa.

"We know that John didn't kiss Lisa."

c. (N) ??Vi vet (at) Lisa kysset Jon ikke.

d. (S) ??Vi vet (att) Lisa kysst Jan inte.

e. (N) ??Vi vet (at) Jon kysset ikke Lisa.

f. (S) ??Vi vet (att) Jan kysst inte Lisa.

(107c.) and (d.) illustrate embedded topicalization of a non-subject; (107e.) and (f.) topicalization of an embedded subject. Note that when a propositional operator of the type which induces SAI in English is topicalized, the resulting embedded INFL2 structure is far more acceptable: this is

illustrated by the Norwegian example in (108) below:

(108) Jeg sa til Jon, at aldri hadde jeg hørt sårt bråk.
I said to John that never had I heard such noise

"I said to John that never had I heard such noise."

Topicalization inside a clause headed by a WH-phrase is generally ungrammatical in Norwegian and Swedish:

(109)a. (N) Vi vet hvem Jon ikke kysset.

b. (S) Vi vet vem Jan inte kysst.
who

"We know who John didn't kiss."

c. (N) *Vi vet hvem Jon kysset ikke.

(S) *Vi vet vem Jan kysst inte.

This is also true of the WH-complementizer <u>om</u>:

(110)a. (N) Jeg spurte om Jon ikke hadde lest den bok.

b. (S) Jag frågade om Jan inte had låst boken.
asked if

"I asked if John hadn't read the book."

c. (N) *Jeg spurte om Jon hadde ikke lest den bok.

(S) *Jag frågade om Jan hade inte l st boken.

In Yiddish, embedded INFL2-related topicalization is far less marked, as illustrated in (111) below (examples are from Travis 1984):

(111)a. Ikh meyn az di froy hot gekoyft dos bukh in der
I think that the woman has bought the book in the

krom.
store

"I think the woman bought the book in the store."

b. Ikh meyn az in der krom hot di froy gekoyft dos bukh.

c. Ikh meyn az dos bukh hot di froy gekoyft in der krom.

All of (111a.-c.) are perfectly grammatical, though obviously they differ in emphasis. It is not the case, however, that INFL2 occurs freely in subordinate clauses in Yiddish: just as in the Scandinavian languages, when a WH phrase is present, no inversion may take place:

(112)a. Ikh veys nit vemen zi hot gezen zuntik.
I know not who she has seen Sunday

"I don't know who she saw on Sunday."

b. *Ikh veys nit vemen hot zi gezen zuntik.

c. *Ikh veys nit vemen zuntik hot zi gezen.

This also applies to the WH-complementizer <u>tsi</u>:

(113)a. Ikh veys nit tsi er hot im gezen in Paris.
I know not if he has him seen in Paris

"I don't know if he saw him in Paris."

b. *Ikh veys nit tsi hot er im gezen in Paris

c. *Ikh veys nit tsi in Paris hot er im gezen.

Now, we have characterized INFL2 in Type I (NCC) languages on the basis of two distinctive properties. First of all, A'-type P-subjects are necessary for all independent AGR-domains, and only independent AGR-domains; secondly, complementizers are specified for AGR-features. The question now arises as to which of these properties differentiates Type I from Type II languages.

We have seen that in Frisian, the second property (AGR-specified complementizers) holds only partially, accounting for "exceptional cases" of INFL2-related topicalization; however, it appears that INFL2-related topicalization in Type II languages is of a rather different nature. In particular, note that Type II languages do not obey the NCC; as in German and Frisian, both Norwegian and Swedish bridge verbs may optionally take finite complements without complementizers, but unlike in Type I languages, INFL2 is not obligatory (and is in fact marked) in such contexts. This is illustrated below:

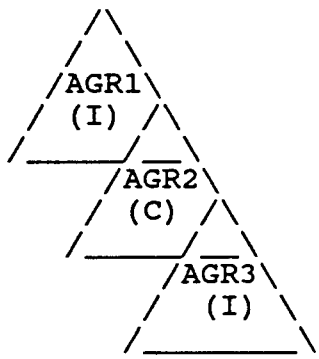
(114)a. (N) Jeg tror (at) hann ikke skulle kommer.
b. (S) Jag tror (att) han inte skulle komma.
I believe (that) he not shall come
"I believe that he won't come."

As can be seen from the placement of <u>ikke/inte</u> in (120), INFL2

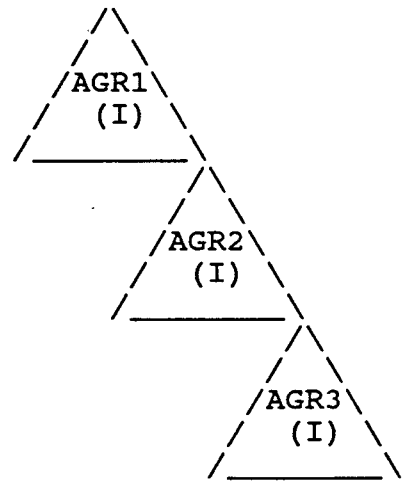
fails to take place in a subordinate clause even without a complementizer. This suggests that the type II languages differ rather more radically from typical Type I languages than does Frisian.

Suppose, as a first step, we were to claim that complementizers in Type II languages are never specified for AGR, in contrast to complementizers in Type I languages. Then the INFL of finite embedded clauses in Type II languages would always contain independent AGR features, since AGR-domains would not overlap. This is illlustrated in (115), where (a.) and (b.) are schematic representation of AGR-domains in Type I and Type II languages, respectively:

(115)a.



b.



(C) here stands for complementizer, (I) for INFL.

In (a.), every AGR-domain except that of AGR1 is dependent; AGR2, which is contained in the complementizer, "links" AGR3 to

AGR1. On the other hand, in (b.) all AGR-domains are independent, since the complementizer contains no AGR-specification and cannot act as a linker.

Now, let us return to (99), repeated here as (116):

(116) Every independent AGR-domain must have a unique A'-type P-subject at S-structure.

Given our assumptions, this predicts that INFL2 must take place in <u>every</u> subordinate clause in Type II languages, since every subordinate AGR-domain must be independent, by the definition in (98). But as we have seen, this is incorrect. It thus appears that only the topmost independent AGR-domain <u>must</u> have an A'-type P-subject; an A'-type P-subject is only optional in other independent AGR domains (with varying degrees of acceptability). In order to incorporate this difference into our model, let us propose the following condition for Type II languages:

(117)a. A root AGR-domain must have a unique A'-type P-subject.

b. Any independent AGR-domain may have an A'-type P-subject.

We define the notion "root AGR-domain" as in (118):

(118) AGR_i is a root AGR iff there is no AGR_j, such that AGR_j

maximally c-commands AGR_i.

Condition (117) will account for most cases of INFL2-related topicalization in Type II languages. One notable set of exceptions to (117b.) remain, however; as we have seen, topicalization is always ungrammatical in non-root clauses headed by a WH-element. One possibility is that some independent semantic constraint might prevent topicalization in WH-complements; in that case, we might expect a contrast in English between embedded topicalization in <u>that</u> clauses and WH-headed clauses, quite independently of SAI. Such a contrast does indeed seem to exist:

(119)a. I remember that George, our grandmother always liked best.

b. ??I don't remember whether George, our grandmother liked best.

However, data from Icelandic, to be reviewed below, show that any semantic constraint must be at least to a certain extent subject to cross-linguistic variation. Accordingly, let us turn to Icelandic, the sole Type III language in our classification.

Icelandic in many ways resembles Type II languages: it has an underlying SVO order, exhibits no NCC effect (since the complementizer <u>ad</u> cannot delete, this is unsurprising) and

allows INFL2-related topicalization in subordinate clauses. However, it differs from Type II languages in two significant ways. First of all, in contrast to the continental Scandinavian languages - see (104) above - INFL2 clauses are the unmarked case in subordinate contexts, as shown in (120) by the (identical) position of sentence adverbials in main and subordinate clauses (all examples from Thrainsson 1986):

(120)a. Helgi hefur trulega keypt bokina.
Helga has (indic.) probably kept the-book

"Helga has probably kept the book."

b. Jon segir aδ Helgi hafi trulega keypt bokina.
John says that Helga has (subj.)

"John says that Helga has probably kept the book."

(Note: δ is used here for the voiced interdental fricative, and th for the voiceless inter-dental fricative).

The lack of an asymmetry between root and non-root clauses suggests that the basic word order of Icelandic is not S INFL VP, as in Type II languages, but INFL S VP. This hypothesis is supported by the fact that so-called "narrative inversion" (NI) sentences, which are declarative sentences with V1 order, are more common in Icelandic than in other Germanic languages; such clauses may also be embedded, as shown in (121):

(121)a. Koma their nu ad storum helli...
Come they now to a-big cave

"Then they get to a big cave."

b. Hann sagði að hefdu their tha komið að storum helli.
He said that have they then come to a-big cave

"He said that they had then come to a big cave."

If the basic word-order of Icelandic is INFL-first, then NI sentences are marked (and generally, anachronistic) examples of clauses in which INFL2-related topicalization has not taken place, perhaps reflecting a stage in the development of the language where INFL2 was not obligatory.

The second distinguishing characteristic of Type III languages is that INFL2-related topicalization is not blocked in clauses headed by WH-elements; as shown by (122b.), this includes INFL1 yes/no question inversion:

(122)a. Jon spurði hvort thessum hring hefði einhver stolið.
John asked whether this ring had anyone stolen

"John asked whether anyone had stolen this ring."

b. Hann spurði hvort hefði komið her i dag einhver
He asked whether had come here today any

postur.
post

"He asked whether any mail had come here today."

Maling (1980) observes that it is easier to front non-subjects when the subject is absent or indefinite: this is a familiar property of topicalization, common to all Germanic languages.

It thus appears that Icelandic differs in two significant

ways from Type II languages. Firstly, INFL2-type clauses are the rule in embedded contexts. Within our framework, we can capture this quite simply by claiming that Type III languages resemble Type I languages in requiring every (independent) AGR-domain to have an A'-type P-subject, but resemble Type II languages in having a complementizer unspecified for AGR. Secondly, unlike either Type I or Type II languages, Type III word order is INFL S VP: in other words, Icelandic is head-initial, the "mirror image" of Type I languages, which are head-final.

There is an interesting corollary to our claim that secondary A'-type predication is obligatory in Icelandic, just as it is in Type I languages. This is that we should expect an A'-type pleonastic to exist in Icelandic, parallel to German es, which will be present whenever there has been no INFL2-related topicalization. In fact, just such an element does exist: expletive thaδ. As noted by Maling and Zaenen (1978), Thrainsson (1979), and others since, thad shows up just in case no other element has been fronted; in that case, it occupies the S-initial XP position. To quote Maling and Zaenen, "The function of thaδ in Icelandic syntax becomes clear: it serves simply to satisfy the verb-second constraint.". Examples of sentences containing thaδ are given below:

(123)a. (Thaδ)* var dansaδ i gaer.

There was danced yesterday

"There was dancing yesterday."

b. I gaer var (*thaδ) dansaδ.

c. (Thaδ)* drekka margir vin a Islandi.
There drink many wine in Iceland

"Many people drink wine in Iceland."

d. A Islandi (*thaδ) drekka margir vin.

e. Hvaδ er (*thaδ) a jörðinni ?
What is there on the-ground

"What is there on the ground ?"

f. Var (*thaδ) mikill snjor a jörðinni i gaer ?
Was (there) much snow on the-ground yesterday

"Was there much snow on the ground yesterday ?"

As shown in (123e.) and (f.), thaδ never appears in WH- or yes/no questions; this behaviour is parallel to that of German es - see (87 - 90) above - and precisely what we would expect if thaδ were an A'-type pleonastic element.

Let us now summarize our investigation of INFL2-related variation in Germanic languages. We have found variation along two parameters: firstly, a complementizer may or may not be specified for AGR; and secondly, obligatory INFL2-related topicalization may take place in every independent AGR-domain, or may be limited to a "root" AGR-domain. This is shown in the cross-classification below:

(124)

	Type I	Type II	Type III
Configuration	S VP INFL	S INFL VP	INFL S VP
AGR-in-Comp	yes	no	no
Obligatory non-root INFL2	yes	no	yes

"Gaps in the paradigm" in (124 indicate that there may be further generalizations underlying those already investigated; in particular, it might be asked why only S VP INFL languages may generate AGR features on the complementizer, and why S INFL VP languages appear to be the only ones not to require an obligatory A'-type P-subject in every independent AGR-domain. We will not attempt to answer these questions here; to do so would require a much larger and more detailed examination, well beyond the bounds of the present study. We gladly leave such an investigation to future research.

Nevertheless, in spite of the limited nature of our own preliminary investigation, it seems to me that the approach taken here to INFL2-related phenomena appears very promising, both in terms of clarifying the relations between the syntax of Germanic languages, and in terms of providing an explanatory framework within which to explore the relevant issues.

In the next sections, we will further expand our cross-linguistic perspective, examining SAI-type processes in Romance languages, beginning with the rule of Complex Inversion which

characterizes most dialects of French as well as several northern Italian dialects.

4.45 Complex Inversion

We now turn to our third case of SAI-type inversion, exemplified by the rule of Complex Inversion (CI) in French and those Italian dialects (notably Trentino and Modenese - see Brandi & Cordin 1981) with subject clitics. Though, as we shall see, CI differs in many ways from the rules we have been considering up to now, it shows most of the criterial properties of SAI-type inversion; it is triggered by a fronted WH-element, involves movement of a tensed $[I,V]^{\circ}$ constituent, and shows a root/non-root asymmetry. The fourth criterion, the subject/non-subject asymmetry, is absent in CI environments for independent reasons, as we shall see below.

Given, then, that CI is an SAI-type of process, we must account for the differences between it and the other rules we have been considering. Let us first, however, illustrate the basic workings of the process. Examples will be from French; other languages with CI act alike, except where noted otherwise.

CI occurs in matrix yes/no and non-subject WH-questions; with certain adverbs (including monotone-decreasing adverbs, as in English, but also peut-être "perhaps", sans doute

"doubtlessly", and ainsi "thus", which do not trigger SAI); (more maginally) with exclamative si preposing (the equivalent of so-preposing in English); with exclamative WH-phrases; and, finally, in certain conditional environments. It does not occur in topicalization structures (which are marginal in French) nor in tags (which are invariant in French, and show none of the polarity properties of tag-questions in English) nor in conjoined structures. The basic triggering environments are illustrated below (example (125f.) is from Goldsmith 1981):

(125)a. A-t-il vu Jean hier ? (yes/no Q)
Has-he seen John yesterday

"Did he see John yesterday ?"

b. Qui a-t-il vu hier ? (WH-Q)
Who

"Who did he see yesterday ?"

c. Peut-être a-t-il vu Jean hier. (adverb)
Perhaps

"Perhaps he saw John yesterday."

d. Si jolie était-elle qu'il l'avait regardée toute la
So pretty was-she that he her-had looked at all the

soirée. (si-preposing)
evening

"She was so pretty he watched her the whole evening."

e. Quel fou était-il ! (WH-exclamative)
What fool was-he

"What a fool he was !"

f. Rentrait-il avant onze heures du soir, elle l'attendait

Returned-he before eleven P.M she him-awaited

dans le vestibule. (conditional)
in the hall

"If he returned before eleven P.M, she would be waiting for him in the hall."

It will be observed that the triggering environments for CI are very much like those for English SAI; apart from WH-exclamatives and some of the adverbs, all CI-triggering elements are also SAI-triggers, and the converse is true with the exception of tags and conjunctive so-preposing, neither of which exist in the same form in French as in English. This striking similarity constitutes a strong argument that the two rules are manifestations of the same syntactic phenomenon. As for the differences, we have already argued that the behaviour of exclamative WH-phrases in English is aberrant; French exclamatives thus represent the unmarked option vis à vis inversion. The difference between CI- and SAI- triggering adverbials appears to be a relatively low-level lexical disparity (though for that very reason interesting, in that the French-English discrepancy constitutes a further argument that there is at least some purely lexical learning involved in acquiring SAI-type rules). As for the failure of CI in conjunctive si-preposing and tag-questions, this is entirely expected, since neither of these English constructions has a precise French equivalent.

We thus conclude that CI-triggers are almost identical to SAI-triggers, and that this constitutes evidence for treating the two processes as closely related. From a cross-linguistic perspective, French resembles English and differs from the other Germanic languages in having a lexically restricted set of possible A'-type P-subjects.

CI differs from all the languages we have been considering up to now in having subject clitics (SCLs). In ordinary tensed sentences, a SCL precedes the tensed verb; where CI applies, this order is reversed. This basic contrast is illustrated in (126) below. Note that both yes/no and (non-echo) WH-questions can be formed in French without CI.

(126)a. Jean a vu Marie hier.
John has seen Mary yesterday

"John saw Mary yesterday."

b. Jean a-t-il vu Marie hier ?

"Did John see Mary yesterday ?"

c. Qui (Jean) a-t-il vu hier ?
Who

"Who did John see yesterday ?"

In standard French (SF), lexical subjects and SCLs are in complementary distribution in clauses where CI does not apply. This complementarity breaks down, however, where CI has taken place; in that case, a SCL is obligatory and a lexical subject

optional. In colloquial French (CF), SCL "doubling" is common and quite unmarked in any tensed clause. SCLs never appear in infinitives or other untensed clauses; since French lacks whatever mechanism accounts for exceptional Case-marking constructions in English, there is a general prohibition against lexical subjects in untensed clauses which independently accounts for this fact. This array of data is exemplified in (127 - 129) below:

(127)a. (Jean*) il est parti. (SF)
John he is left

b. Jean (il*) est parti. (SF)

c. (Jean) il est parti. (CF)

d. Jean (il) est parti. (CF)

e. *(Jean il) est parti. (CF & SF)

"John left."

(128) (Jean) est-il parti ? (CF & SF)
John is-he left

"Did John leave ?"

(129)a. Je veux qu'il parte.
I want that he leave

"I want him to leave."

b. Il veut (il*) partir.
He wants he to-leave

"He wants to leave."

Finally, if a subject is WH-extracted in French, then no SCL

can appear in INFL, as illustrated by (130) below:

(130) Qui est(-il*) arrivé hier ?
Who is(-he) arrived yesterday

"Who arrived yesterday ?"

There is much (as yet unresolved) controversy over where to generate SCLs in French, with at least three positions having been suggested in the literature: Kayne (1975) and Rizzi (1984), amongst many others, argue that the SCL is in external argument position, Sportiche (1983) and Roberge (1986) that it is generated in INFL, and Morin (1979) and Safir (1982) that it is an affix on the verb. All of these analyses have advantages and drawbacks, which we will very briefly review here; for more extensive accounts, see the references above, as well as Jaeggli (1982) and Safir (1985).

The major advantage of the "external-argument" analysis is that it accounts straightforwardly for the complementary distribution of lexical subjects and SCLs in declarative sentences in SF, since they occupy the same position at D-structure; however, where complementary distribution breaks down (i.e., in the "SCL-doubling" construction of CF and various Italian dialects, as well as in CI environments in general) this analysis encounters obvious problems. These problems are most acute for dialects such as the Trentino dialect of Italian, where a (third person) SCL is obligatory,

whether or not a lexical NP is present. This is illustrated in the following paradigm (from Safir 1985):

(131)a. El Mario *(el) magna.
Det. Mario SCL eats

"Mario eats."

b. *(El) magna.

"He eats."

The usual solution to this problem (offered, for example, by Kayne 1984), is to propose that the lexical subject NP occupies a base-generated S-external adjoined position just in case it cooccurs with a SCL; however, without independent motivation, such a solution appears at best ad-hoc.

The "verbal-affix" analysis of SCLs encounters different kinds of problems. In particular, it predicts that, ceteris paribus, SCLs should pattern in a similar manner to object clitics (OCLs), which are standardly analyzed as base-generated verbal affixes. But as first pointed out by Kayne (1972), there are a number of systematic differences between OCLs and SCLs with respect to ordering, the position of the negative scope particle ne, and the relative "independence" of the SCL in sentences such as that in (132):

(132)a. ?Qui ca, ils ?
Who that they(SCL)

"Who's that, them ?"

b. *Qui ca, les ?
them(OCL)

The verbal affix analysis has no way to account for any of these differences, nor indeed, for the different relative ordering of the SCL and the verb in declarative and CI environments (Safir 1982, for example, is forced to claim that the SCL "can be realized on either side of the verb, but only once", a strange claim in view of the fact that clitics are in general strictly ordered both with respect to each other and the verb to which they are attached). In view of these deficiencies, we conclude that the verbal-affix analysis must be rejected.

Let us then turn to the "AGR" analysis of SCLs, discussed most extensively by Roberge (1986), who proposes that SCLs are generated in INFL as an instantiation of AGR itself. This is clearly more satisfactory for SCL-doubling cases; in Roberge's analysis, a simple "Case-transmission parameter" distinguishes languages with SCL-doubling (where the SCL may transmit Case to the [NP,S] position) from those without (where the SCL absorbs and may not transmit nominative Case). However, the assumption that a SCL is an instantiation of AGR runs into certain problems; though SCLs resemble the "rich" AGR of standard Italian or Spanish in that both constitute "identifiers" for pro (the governed empty pronominal proposed by Chomsky 1982

as the [NP,S] subject in "pro-drop" languages), it is hard to account for why subject-verb agreement morphology should surface in SCL languages in exactly the same way as in non-SCL languages without the assumption that AGR and the SCL are separate entities. Moreover, there is evidence that the SCL in fact occupies a different underlying position from INFL (and hence, from AGR). In SF infinitivals, the verb follows both the negative particle pas and the "scope-marking" particle ne; this is shown below:

(133)a. J'ai essayé de ne pas les déranger.
I(SCL)-have tried to SCOPE NEG them(OCL)-disturb

"I tried not to disturb them."

b. *J'ai essayé de ne les déranger pas.

On the other hand, in finite clauses, the (tensed) verb generally occurs in between the scope-marker and the NEG particle, as shown in (134):

(134)a. Je sais qu'ils ne dorment pas.
I(SCL) know that-they(SCL) SCOPE sleep NEG

"I know that they aren't asleep."

b. *Je sais qu'ils ne pas dorment.

This word-order difference has been plausibly linked to the rule of V-raising (see Emonds 1978, Pollock 1987, and Chapter 3 of this thesis) which we have argued involves the movement of a

verb into tensed INFL. If so, then INFL occupies a position preceding NEG (which is presumably in VP-adjoined position, as in English and Germanic) but following the scope-marker ne. However, a SCL always precedes ne, as shown by (135):

(135)a. Il n'est pas arrivé encore.
He(SCL) SCOPE-is NEG arrived yet

"He hasn't arrived yet."

b. *N'il est pas arrivé.

It therefore appears that INFL and the SCL are in different positions. One possible way round this (suggested by Rizzi 1984) is to claim that both the SCL and ne are originally in INFL, which will then, subsequent to V-raising, contain three elements: ne, the SCL, and the raised tensed verb, which will have adjoined to the right of INFL. However, if so, we might expect the tripartite INFL constituent to act as a unit for the purposes of CI; this is not the case, as can be seen from (136):

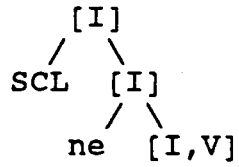
(136)a. Ne m'a-t-il pas vu hier ?
SCOPE me(OCL)-has-he(SCL) NEG seen yesterday

"Didn't he see me yesterday ?"

Here ne and the verbal complex act as a separate unit from the SCL. We are forced to the conclusion that either the SCL has been lowered into a post-INFL position, or the (ne + V)

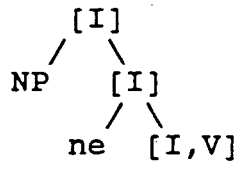
combination has been raised into a pre-SCL position; in either case, the SCL must be a separate constituent. If so, then we must adopt a more elaborated structure for the "INFL-complex" such as that in (137):

(137)



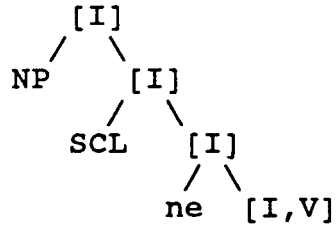
Now, note that in a theory such as ours, in which X-bar principles are derivative rather than primitive, the representation in (137) is in fact isomorphic with a structure containing a lexical subject NP:

(138)



In other words, without further stipulation, we are able to account for the structural parallels between lexical subjects and SCLs. Now, let us take a SCL-doubling structure:

(139)



Here, the lexical NP is automatically defined as "S-adjoined", once again, without further stipulation. Note that we can still account for doubling/non-doubling variation by invoking Roberge's Case-transmission parameter, since we can still uniquely identify the lexical "subject" NP as [NP, [IP]] (i.e., [NP,S]) in (139). We are thus able to combine the advantages of the "external argument" and "AGR" analyses of SCLs, without at the same time incorporating their drawbacks.

One residual problem remains, however. Roberge points out that the status of a lexical subject in SCL-doubling constructions in CF is rather different from that of the optional lexical subject in CI environments: whereas the former is subject to an "indefiniteness constraint" preventing a quantified or indefinite NP from appearing in a doubled position, the latter is not. This is illustrated in (140) below:

(140)a. Mon oncle il est sorti par la fenêtre.
My uncle he(SCL) is gone out by the window

"My uncle left by the window."

b. Mon oncle est-il sorti par la fenêtre ?

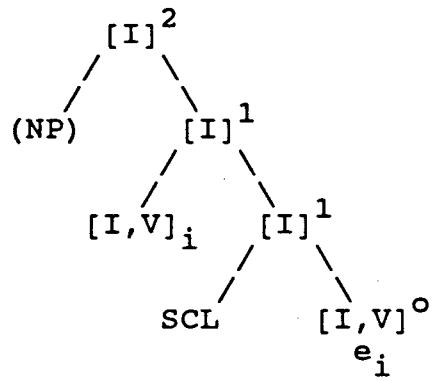
c. *Un homme il est sorti par la fenêtre.
A man

"A man left by the window."

d. Un homme est-il sorti par la fenêtre ?

Roberge handles this contrast by assuming that in SCL-doubling constructions the doubled NP is in the [NP, S] position but in CI constructions it is in an S-external position. This explanation is not available, however, in a theory of phrase-structure such as ours: the doubled NP cannot be structurally distinguished in the two constructions. Nevertheless, we can account for the observed contrast quite easily through a slight adjustment to the Case-transmission parameter. Let us assume the following derived structure for CI:

(141)



In essence, we have analyzed CI as an adjunction of $[I,V]^{\circ}$ to $[I]^{1}$ (see Jaeggli 1982 for a similar approach); we shall motivate the rule in more detail immediately below. Now, notice that following CI, $[I,V]^{\circ}$ is directly adjacent to the "external" lexical NP, whereas in non-CI SCL-doubling constructions - see (139) above - the SCL intervenes between [I,V] (which carries nominative Case-assigning AGR features)

and the external NP. Assuming an adjacency condition on Case-assignment, it is now easy to see why nominative Case can be freely assigned to the "external" NP in (141), but not in (139); only in the former case is there strict adjacency between the Case-assigner and the NP to which Case has been assigned.

Note that this analysis implies that Case can be assigned both at NP-structure and at S-structure: AGR will assign nominative to the SCL at NP-structure, then, following CI, it will assign it again to the external NP at S-structure. Obviously, "double Case-marking" of this type is an extremely limited phenomenon; in fact, I would like to claim that it is limited to structures involving clitics. Suppose we were to say that SCLs had the unique ability not to absorb Case; that is, while bearing Case-morphology, they do not deprive the AGR Case-assigner of its Case-assigning features. This would successfully account for the contrast between CI and SCL-doubling; in the latter case, the doubled NP would receive Case only through transmission via a SCL, whereas in the former, the external NP would receive nominative Case directly under government by AGR.

While this might appear ad-hoc, note that there is independent support for the idea that Case must be assigned at S-structure in French. This evidence is supplied by a well-known contrast first pointed out by Kayne (1984), and

reproduced below:

(142)a. *Je crois Jean être plus intelligent que moi.
I(SCL) believe John to-be more intelligent than me

"I believe John to be more intelligent than me."

b. Qui crois-tu être plus intelligent que moi ?
Who you(SCL)

"Who do you believe to be more intelligent than me ?"

(142b.) shows that Case-assignment to a WH-trace in COMP is possible in French, whereas Case-marking of a subject NP is blocked in ECM structures such as (142a.). Now, since the intermediate WH-trace in COMP is not present at NP-structure, Case-assignment must take place at S-structure to license a structure such as (142b.), providing independent evidence for the possibility of S-structure Case-assignment.

Before continuing, let us briefly turn to a further parametric variation in Case-assignment in CI languages. It turns out that in Italian dialects with SCLs, the distribution of lexical NPs in CI is different from either SF or CF, as illustrated below for Trentino (other dialects appear to behave identically, as far as I can ascertain) :

(143)a. Quando magnelo ?
When eats-he(SCL)

"When does he eat ?"

b. *Quando magna ?

eats

c. *Quando el Mario magnelo ?

As can be seen from (143c.), a lexical NP in CI environments is prohibited in Trentino, whereas in declarative environments with SCLs it is optional - see (133) above. (Apparently, this prohibition is less strong in yes/no questions with CI than in WH-questions; I have no explanation as to why this should be the case). We thus get the following distribution of co-occurrence restrictions between SCLs and lexical NPs in SF, CF, and Trentino:

(144)

	[-CI,+SCL]	[+CI,+SCL]
SF	*NP	(NP)
CF	(NP)	(NP)
Trentino	(NP)	*NP

Now, we have ascribed the difference between SF and CF to a Case-transmission parameter: Nominative Case can be either assigned directly to a lexical NP, under adjacency, or by transmission via a SCL; in French, the latter possibility is only instantiated in CF. In CI cases, on the other hand, AGR, the Case-assigning element, is adjacent to a lexical NP, and may assign it Case directly, accounting for the optional presence of a lexical NP in both SF and CF. Now, obviously, Trentino is also a Case-transmission language - indeed, it

appears that direct Case-assignment to a preverbal subject is impossible, as witnessed by the ungrammaticality of (133) above. But in that case, we have an immediate explanation for the ungrammaticality of (143c.) as well; since there is no direct nominative Case-assignment, a preverbal lexical NP will never appear unless transmitted Case by a SCL - but in CI environments, transmission cannot take place, since the SCL and the lexical NP are separated by the inflected verb carrying AGR features.

We thus account for some of the peculiarities of "SCL-doubling" in CI languages by assuming two Case-assigning mechanisms - direct Case-asssignment, and SCL Case-transmission. SF employs only the former, Trentino only the latter, and CF both. We have also proposed that direct Case-assignment be allowed to occur - in a limited number of cases - at S-structure.

Let us now return to our investigation of CI as an SAI-type rule - that is, as an instance of a predication-related syntactic process.

First of all, let us make the following assumptions:

(a) A SCL counts as an A-type P-subject.

(b) Tensed [I,V] contains AGR-features.

Together with the conditions on predication in (39.III), repeated below in a slightly modified version as (145), (a) and (b) provide an immediate explanation for CI:

(145) II. Every AGR-domain must have a unique X-type P-subject at level n.

III. Every predication subject must be AGR-linked.

At NP-structure, a SCL will count as the A-type P-subject for the AGR-domain headed by AGR-features in [I,V]; therefore, by (145.II), a lexical NP in the [NP,S] position will not count as a P-subject. At S-structure, if an inversion-inducing operator is fronted, it will become the (A'-type) P-subject for the same AGR-domain; therefore, it must be linked to AGR by (145.III). Now, the A-type P-subject SCL will be "opaque" for AGR-linking, by (40), repeated below as (146):

(146) AGR-linking: A P-subject α is linked to AGR_i iff α minimally governs AGR_i.

Therefore, CI takes place in order to link AGR to its A'-type P-subject at S-structure, thus satisfying (145.II).

However, so far we have only provided half an explanation for CI in French. In particular, whereas we have accounted for the grammaticality of CI structures, we have not accounted for why a sentence such as (147) below should be ungrammatical:

(147) *Où est Jean allé ?
Where is gone

"Where did John go ?"

(147) has undergone an inversion parallel to English SAI or INFL2. Now, we have seen that the same principles which motivate SAI-type inversion rules in English and Germanic rules apply to French; and given that lexical external argument NPs count as A-type P-subjects in French when not accompanied by a SCL, it is hard to see why (147) should be impossible.

In order to explain the ungrammaticality of (147), we will appeal once more to Case-theory, this time in the form of the statement in (148):

(148) In SCL languages, nominative Case is assigned to SCLs at NP-structure, but to lexical NPs at S-structure.

(148) is an extension of the claim we have already made that S-structure Case-assignment is a peculiar property of SCL languages. It is easy to see how (148) rules out SAI-type inversions in CI languages: assuming nominative Case is assigned directionally (i.e., leftward), a structure such as that in (147) contains a subject NP in a position where it can neither be transmitted Case by a SCL nor assigned Case directly by AGR; hence, it fails the Case-filter, and is ruled out.

(148) implies that no language with CI will also have an SAI-type rule: as far as I know, there are no counter-examples to this claim.

We have now accounted for the main characteristics of CI; there remain, however, the following oustanding issues:

(a) The subject/non-subject asymmetry associated with CI.
(b) The root/non-root asymmetry associated with CI.
(c) The relation between CI and Stylistic Inversion (SI).

Let us, accordingly, deal with these topics one at a time.

First of all, let us deal with the subject/non-subject asymmetry exemplified in (128) above, and in (149) below:

(149)a. Qui veut lutter pour un monde plus juste ?
Who wants to-struggle for a world more fair

"Who wants to struggle for a fairer world"

b.*Qui veut-il lutter pour un monde plus juste ?
SCL

As can be seen from the examples above, inversion never takes place with subject extraction in CI languages, which in this respect resemble English and differ from the INFL2 Germanic languages. We have claimed that the English/Germanic contrast is a consequence of optional versus obligatory secondary (A'-type) predication; since French and other CI dialects pattern

with English SAI in allowing inversion only with a restricted set of syntactic operators, it seems natural to treat them, like English, as languages with optional secondary predication. Now, recall that in English, we claimed that an operator in cases of "subject-extraction" could satisfy both A-type and A'-type predication without the need for SAI. The same is true for CI languages, in that a WH-phrase in [NP,S] position may be linked to AGR at both NP-structure (for A-type predication) and S-structure (for A'-type predication), satisfying predication without the need for CI.

So far,we have shown that CI need not take place, but not why it cannot take place. In other words, though we have an explanation for the grammaticality of sentences such as (149a.), we still need an explanation for the ungrammaticality of sentences such as (149b.). Note that we cannot appeal to a general constraint against extraction out of the [NP,S] subject position of clauses containing a SCL, as proposed by Jaeggli (1984); Roberge (1986) shows firstly that such a constraint must be parametrized, and secondly that it functions differently in "true" SCL-doubling cases and in CI cases, as discussed above. Instead, it appears that we must once again appeal to (a strong version of) the last resort theory of movement rules, in which the application of a syntactic movement rule will be ungrammatical unless forced by some grammatical principle(s).

Let us then turn to the second asymmetry connected with both CI and SAI: the root/non-root asymmetry. Just as in English SAI, CI is generally ungrammatical in subordinate contexts, as illustrated in (150) below:

(150)a. (*Je ne sais pas si) Jean est-il revenu chez lui.
I know not if John is-he returned his place
"(*I don't know whether) did John go home."

b. (*Je me demande) ou est-il allé.
I me ask where is-he gone
"(*I wonder) where has he gone."

We have already discussed in detail (in 4.44) parametric variation between the INFL2 languages with respect to this asymmetry; there, we distinguished between Type I languages (which showed NCC effects) Type II languages (which showed root/non-root effects) and Type III languages (which showed neither). The question now arises as to which category CI languages fall into.

One clue is provided by major constitutent order; French has the order S INFL VP, like Type II languages and unlike Type I or Type III languages. Note also that French shows a *[COMP t] effect, like all type II languages, but unlike Type I or Type III languages: the specially indexed complementizer qui must be present in subject extractions from tensed clauses, as shown in (151c.) below:

(151)a. Qui penses-tu que Jean a vu hier ?
Who think you(SCL) that John has seen yesterday

"Who do you think John saw yesterday ?"

b. *Qui penses-tu que est parti hier ?
is left

c. Qui penses-tu qui est parti hier ?
who

We might then speculate that French ought to display other typical Type II characteristics - marked inversion in subordinate clauses, for example. Unfortunately, French topicalization - marginal even in matrix clauses - is generally ungrammatical in subordinate contexts, and extremely difficult to distinguish from left dislocation on the one hand and SCL-doubling on the other. For example, while it is possible to prepose certain CI-inducing adverbs in embedded contexts, as shown in (152) below, and it is true that CI fails to take place in such contexts, adverb preposing in clauses headed by a complementizer is always accompanied by a pause, suggesting that left-dislocation rather than topicalization may be responsible:

(152)a. Peutêtre Jean va-t-il au cinéma ?
Perhaps John goes-he(SCL) to-the cinema

"Perhaps John is going to the movies."

b. *Je me demande si peutêtre Jean va-t-il au cinéma.
I(SCL) wonder

c. Je me demande si, peutêtre, Jean va au cinéma.

"I wonder whether perhaps John is going to the movies."

Embedded topicalization thus provides no reliable relevant data. Moreover, complementizers are undeletable in French; this deprives us of another criterion for identifying Type II languages, which show no NCC effect in subordinate clauses without a lexical complementizer.

Therefore, in view of the lack of strong evidence available, we can only tentatively class French as a Type II language with respect to the root/non-root asymmetry.

One possible source of counterevidence is provided by Goldsmith (1981), who claims that NCC effects do in fact occur in French. Goldsmith points out that in many varieties of French CI alternates in matrix clauses with the presence of an overt (lexical) complementizer. In SF this alternation is limited to certain adverbial environments, such as those in (153) below :

(153)a. Peut-être est-il allé au cinéma.
Perhaps is-he gone to-the movies

b. Peut-être qu'il est allé au cinéma.
that-he is

"Perhaps he went to the movies."

However, in some varieties of CF, particularly the variety of

Québécois spoken in and around Montréal (QF), the alternation is much more widespread; any WH-question which would require CI in SF may be realized in QF with an overt complementizer instead:

(154)a. Comment vas-tu ? (SF)
How go-you(SCL)

b. Comment que tu vas ? (QF)

"How are you going ?"

c. Où Jean va-t-il ? (SF)
Where goes-he(SCL)

d. Où que Jean va ? (QF)

"Where is John going ?"

Note that even when no CI-inducing element is present, it is marginally possible in both CF anmd SF to have a lexical complementizer in (non-echo) matrix questions, exclamatives, and imperatives:

(155)a. Si tu vas ? (CF)
If you(SCL) go

"Are you going ?"

b. Qu'elle est belle ! (SF)
That-she(SCL) is beautiful

"How beautiful she is !"

c. Qu'ils meurent ! (SF)
That-they(SCL) die

"Let them die !"

Again, this is generally impossible in English, though it appears that some Germanic INFL2 languages tolerate similar "embedded" constructions in root contexts: see Haider (1986b.).

Now, while this data shows that, at least in some French dialects, clauses headed by complementizers are not confined to embedded contexts, it does not show that French obeys the NCC, as opposed to a root/non-root distinction such as characterizes other Type II languages. Evidence of the latter type could only be provided by a direct alternation between lexical complementizers and CI, such as that provided by embedded complementizer-less clauses in German. However, no such alternation exists in French, since, as observed above, complementizers never delete. Rather, it appears that French tolerates two sets of matrix clauses, which we might label "independent" (without a complementizer) and "dependent" (with a complementizer). This distinction is itself to be distinguished from the "dependent/independent AGR parameter" which we introduced in our discussion of INFL2 structures; we assume that complementizers in French, as in Type II Germanic languages, can never be specified for AGR features. While the dependent clause/independent clause distinction might appear ad-hoc, it is well-established in languages as diverse as Tsimshian (Kathy Hunt, p.c.) and Igbo (Goldsmith 1981), both of which allow two distinct sets of matrix clauses, "independent" clauses which carry declarative interpretation and contain no

overt material in a presentential "COMP" position, and "dependent" clauses which carry non-declarative (interrogative, modalized, or negated) interpretation and contain base-generated overt material (including complementizers, question particles, and negative and modal operators) in presentential positions. There appears to be a separate parameter (or parameters) governing the acceptability of dependent clauses in matrix environments: in French, Igbo, and Tsimshian, they are - to varying degrees - acceptable, whereas in English and German they are more marginal.

To conclude, much of the evidence concerning the root/non-root asymmetry in French CI is inconclusive. Mainly on the basis of typological considerations, we have identified French as a Type II (root/non-root) language, as opposed to a Type I (NCC) language. We have also identified a separate parameter, which we might call the "matrix dependent clause parameter", responsible for distinguishing languages which tolerate dependent clauses in matrix environments (with a restricted set of interpretations) from those which as a rule do not.

Let us now turn to our third question: what is the relation between CI and the rule of Stylistic Inversion (SI) in French ? In order to provide an answer to this question, let us first take a look at some of the basic properties of the SI construction.

It is well-known that in SF (though not in CF, where SI is

at best marginal) CI alternates with SI. The latter rule, which in general has a marked, "literary" status, involves a process of "subject-postposing" similar to, but not identical with, the process undergone by post-verbal subjects in the Free Inversion (FI) languages. Its principal characteristics are illustrated in (156) below:

(156)a. Où est(-il*) allé Jean ?
Where is (SCL) gone John

"Where did John go ?"

b. *(Si) est allé Jean ?
If

"Did John go ?"

c. Je ne sais pas combien d'artichauts a mangé Jean.
I know not how-many of-artichokes has eaten Jean

"I don't know how many artichokes John has eaten."

d. Voici le livre que Jean dit qu'aime Marie.
Here-is the book that says that-likes

"Here is the book that John says that Mary likes."

(156a.) shows that SI and CI are mutually incompatible; (156b.), that SI must be licensed by a WH-phrase (or its trace, as in (156d.); (156c.-d.), that there is no root/non-root asymmetry with respect to SI. Thus it can be seen that SI shares some of the properties associated with the SAI-type rules we have been examining - in particular, it is associated with a WH-"trigger" - but in other ways is atypical, showing no

root/non-root asymmetry, for example.

Note that SI differs strongly from the superficially similar process of inversion exemplified in (157):

(157)a. Il est arrivé une femme.
It is arrived a woman

"There arrived a woman."

b. *Il est arrivé Jean.

"There arrived John."

The process in (157) (which we will call "Indefinite Inversion" (II)) requires an obligatory impersonal SCL and, just as in the equivalent English construction with <u>there</u>, is subject to the "Definiteness Effect" extensively discussed in Safir (1985): only an indefinite NP may occupy the post-verbal position. As can be seen from (157a.) there need be no trigger WH-phrase.

We do not wish to go into a detailed analysis of all aspects of the SI construction here; it has been extensively investigated elsewhere (see Kayne & Pollock 1978, Safir 1982, Kayne 1984, Obenauer 1985, to name a few). We do, however, wish to provide an account of the "SAI-type" properties of SI, since these relate directly to our predication-based theory of inversion; in particular, since CI is unnecessary - in fact, ungrammatical - in sentences which have undergone SI, it appears that SI must somehow satisfy the predication conditions

in (145). It is to the question of exactly how SI does satisfy these conditions that we now turn.

First of all, we need to know more precisely what position the "postposed" subject NP in SI occupies. There are two possibilities: the first (advocated in Safir 1982) is that the postposed NP is adjoined to S; the second (assumed in Kayne 1984), that the postposed NP is dominated by an instance of VP. Empirically, there appears little to choose between these structures; conceptually, there are advantages and disadvantages to both. The S-adjunction analysis has the advantage that the postposed subject c-commands its trace, avoiding the need for some special mechanism to coindex the trace in [NP,S] position with its "antecedent" in VP; on the other hand, the VP-adjoined analysis captures the relation between SI and FI in Romance languages by assigning the two processes the same structural description. We will adopt a variant of the latter position, but with the following additional assumptions:

(a) In SI structures, the post-verbal NP is not base-generated, but the product of a syntactic movement rule, possibly a case of "focus-movement" (see Rochemont 1986).
(b) Both nominative Case and the external θ-role are assigned to a <u>preverbal</u> subject NP at NP-structure.
(c) At S-structure, a fronted WH-phrase licenses a "parasitic"

empty category in preverbal subject position (see Kayne 1984).

(d) At LF, the post-verbal subject moves to a presentential position where it is "absorbed" by the previously fronted WH-phrase (see Higginbotham and May 1981, Safir 1985).

It can be seen that though the postposed NPs in SI and in FI structures occupy the same position, we are treating the two processes as separate. This is justified by certain obvious differences between them: in particular, SI is ungrammatical without a triggering operator, unlike FI.

Assumptions (a.-d.) allow us to provide an account for the principle characteristics of SI. At NP-structure, an NP in preverbal subject position receives the external θ-role from VP and nominative Case from (raised) [I,V]. In the mapping from NP-structure to S-structure, this preverbal subject NP adjoins to VP, where it is marked as [+FOCUS]; at the same time, a WH-phrase is fronted to presentential position. At S-structure, the fronted WH-phrase acts as antecedent governor for the "parasitic" (Case-marked) trace in preverbal subject position, as well as for its own trace, thus avoiding an ECP violation (see Kayne 1984 for details). At LF, the focussed NP in postverbal position is raised to presentential position, where it is "absorbed" by the WH-phrase. The "binary quantifier" created by absorption binds both variables, thus satisfying LF

constraints on binding and quantification, and avoiding a binding theory violation. In this way, we account for SI without violating any of the major principles of θ-theory, Case-theory, and binding theory, or the ECP.

Now, let us return to the question of how SI satisfies the predication conditions in (145). At NP-structure, there is no problem; the preverbal subject NP counts as an A-type P-subject for the AGR-domain headed by [I,V], to which it is properly linked.

At S-structure, however, things are less clear. We know that the WH-"trigger" of SI acts as an A'-type P-subject when fronted at S-structure; but we also know that at S-structure the empty category which occupies the [NP, [S]] position cannot be a P-subject - if it were, then AGR-linking could not take place, and SI structures would be ruled out. This means in effect that the variable in preverbal position cannot count as a P-subject. This is an odd conclusion, though one also reached - from a rather different perspective - by Safir (1982); it is all the more odd if we compare the behaviour of the subject trace in SI to that of the intermediate [±WH] traces in German embedded INFL2 structures, which apparently do satisfy A'-type predication (see the examples in (87) above). One possible explanation for the contrast is that A'-type predication is actually successive cyclic; in that case, no trace would count as a P-subject, and SI would be quite unexceptional in this

respect. We will return to this possibility in 4.46 below, when we examine the successive cyclic analysis of V-preposing in Spanish proposed by Torrego (1984).

Finally, let us briefly return to the lack of a root/non-root asymmetry in SI, exemplified in (156) above. Here we have relatively little to say; noting that SI only indirectly satisfies A'-type predication, and that it is only optional in embedded contexts, we might expect it to apply quite freely, unless pre-empted by CI; this indeed appears to be the case (see Kayne and Pollock 1978 for a more detailed examination).

This concludes our discussion of SI, and indeed, of CI; we hope to have shown that both processes are amenable to the predication-based treatment suggested here, which allows a unified approach to the various inversion rules we have been examining in this chapter. In the next section, we will move on to an examination of our final SAI-type inversion process, which characterizes the grammars of romance languages such as Spanish, Portugese, Catalan, standard Italian, and Rumanian.

4.46 Obligatory Inversion in Romance

We have seen that our predication-based account of SAI in English can be extended quite successfully to INFL2 in Germanic and to the processes of CI and SI in French; we will now turn our attention to a similar inversion process which

characterizes questions and related structures in Romance languages. Much of our discussion will centre on Spanish; unless otherwise noted, SAI-type inversion in the other Romance languages shows identical properties.

At the outset, it should be pointed out that the effects of predication-related inversion in Spanish are much harder to detect than SAI, CI, or INFL2 effects. This is due to the operation of FI in Spanish (and in Catalan and Italian), which allows subjects to appear freely in post-verbal position, as illustrated below:

(158)a. Juan contestó la pregunta.
John answered the question

"John answered the question."

b. Contestó la pregunta Juan.

(158a.) and (b.) are identical in meaning. The post-verbal NP in (158b.) is generally assumed to be adjoined to VP from where the verb may properly govern it. Rizzi (1982) shows that the post-verbal NP must be properly governed, since extraction is possible from the post-verbal subject position across a lexical complementizer (thus voiding the *[COMP t] effect), though not from the preverbal subject position.

Another independent property of Romance languages, the Null Subject Parameter (NSP) (briefly discussed in Chapter 3), further confuses the issue. The NSP allows subjects in so-

called "pro-drop" languages to surface without phonetic content, as in (159) below:

(159) Contestó la pregunta.
Answered 3.sing. the question

"He/she/it answered the question."

Together, the NSP and the FIP often make it impossible to tell whether SAI-type inversion has actually taken place in Spanish. For example, a question such as

(160) Qué querian ?
What want 3.pl.

"What do they want ?"

gives us no clue as to whether inversion has taken place or not, whereas a question such as

(161) Qué querian esos dos ?
What want those two

"What do those two want ?"

could either be derived by SAI-type inversion or by Free Inversion. In fact, the only sure way to tell whether SAI-type inversion has taken place in Spanish is to note the ungrammaticality of sentences such as

(162) *Qué esos dos querian ?

Since Free Inversion is generally optional, we conclude that it is the failure of obligatory SAI-type inversion to apply here which leads to ungrammaticality. This is particularly interesting from a learnability standpoint since in order to deduce this from the available data, a child would have to employ negative evidence, which is of course generally considered to be unavailable to the first language learner.

Nevertheless, as argued by Torrego (1984), from which much of the data in this section is taken, we can detect the effects of an obligatory SAI-type inversion rule in Spanish with the following properties:

(a) It is obligatory when triggered by topicalized constituents, and by certain preposed WH-phrases and their traces.
(b) It operates indiscriminately in matrix and subordinate clauses.
(c) It does not operate in infinitival clauses.

Recalling our criteria for the identification of an SAI-type inversion rule, we see that the process under discussion shows at least two diagnostic properties: it is triggered by a preposed WH-phrase and is confined to clauses containing AGR features (as evidenced by its failure to occur in

infinitivals). A third criterion, the root/non-root asymmetry, is not met: SAI-type inversion in Spanish takes place in non-root environments. However, as we have already seen in our discussion of variation in INFL2 languages, the root/non-root asymmetry is at best a weak test for an SAI-type rule, since it is subject to extensive parametric variation. We thus conclude that, at least on an initial inspection, the grammar of Spanish does indeed contain an SAI-type inversion rule, which we will call Obligatory Inversion (OI) to distinguish it from FI. Let us then proceed to a more detailed examination.

First of all, let us examine the elements which trigger OI. Torrego states that inversion is obligatory with argument WH-phrases; this includes non-subcategorized WH PPs, however, as shown by (163) below (=Torrego's (3)):

(163)a. Con quién vendrá Juan hoy ?
With who will come John today

"Who will John come with today ?"

b. *Con quién Juan vendrá hoy ?

At least some (S.American) Spanish speakers however accept (163b.) as grammatical, while still regarding (162) above as ungrammatical. Other WH-phrases which do not trigger obligatory inversion are en que medida "in what way", como "how", cuando "when", por que "why"; in addition, the WH-complementizer si never triggers obligatory inversion.

At first glance, then, Obligatory Inversion in Spanish appears to be even more restricted than SAI in English. However, it turns out that topicalization in Spanish also requires OI, as in INFL2 languages:

(164)a. Un viaje a las Canarias hizo Antonio este verano.
A trip to the Canaries made Antonio this summer

"Antonio made a trip to the Canary Islands this summer."

b. *Un viaje a las Canarias Antonio hizo este verano.

c. Este verano hizo Antonio un viaje a las Canarias.

d. *Este verano Antonio hizo un viaje a las Canarias.

e. Antonio hizo un viaje a las Canarias este verano.

Moreover, exclamative WH-phrases pattern with interrogative WH-phrases in requiring inversion in Spanish, unlike in English:

(165)a. Qué sed me contó Juan que había pasado la gente
How thirsty me-told John that had been the people

aquel dia !
that day

"How thirsty John told me the people were that day !"

b. *Qué sed Juan me contó Juan que había pasado la gente

aquel dia !

Spanish OI thus appears to show some of the restricted characteristics of English SAI and French CI and some of the more general characteristics of INFL2. This appears to provide

evidence that "operator-related" and "topic-related" inversion are to a certain degree independent: a language may show lexical restrictions on the former while still permitting the latter. Note that Spanish OI does not provide counter-evidence to the Designation Convention, which we have argued differentiates rules applying to closed class and open class items; such counter-evidence could only be provided by a rule that applied to a lexically restricted set of open-class items. In terms of our predication-based approach to inversion, Spanish appears to pattern with English and French in allowing optional A'-type predication in both main and subordinate clauses, but with the INFL2 languages in allowing any XP to count as an A'-type P-subject.

Let us then turn from the triggers of OI to its targets. To start with, Torrego shows that OI is restricted to tensed clauses. Her evidence is based on the fact that it is impossible to extract a complement from a WH-island in Spanish just in case V-preposing has taken place (however, see Rivero 1978, 1980, for a different analysis, based on different grammaticality judgements). Extraction of a complement out of a WH-infinitival is, however, always grammatical, as exemplified in (166) below (= Torrego's (60)):

(166)a. *Cuál de estos dices que no sabes a quién regaló
Which of these say-2s.that not know-2s.to whom gave

Juan ?
John

"Which of these do you say that you don't know to whom John gave ?"

b. Cuál de estos dices que no sabes a quién regalarle ?
to give

"Which of these do you not know who to give to ?"

Since if OI had taken place in (166b.) the resulting sentence would have been ungrammatical, it is concluded that no inversion takes place with infinitivals. The fact that OI does not take place in infinitivals is strong evidence that it is "INFL-related" (more strictly speaking, AGR-related) rather than simply V-related, pace Emonds (1985).

Spanish OI also resembles INFL2 in Germanic and CI and SI in French in allowing any tensed verb to occupy "second position". However, with the auxiliaries ser and haber both the finite form of the auxiliary and the main verb must precede the subject in OI contexts. This is illustrated below (examples again from Torrego):

(167)a. Con quién podrá Juan ir a Nueva York ?
With whom will be able John to go to New York

"Who will John be able to go to New York with ?"

b. *Con quién podrá ir Juan a Nueva York ?

c. Qué ha organizado la gente ?
What has organized the people

"What have people organized ?"

d. *Qué ha la gente organizado ?

Now, whereas Torrego analyzes OI as the movement of a tensed verb to clause-initial position (hence her term "V-preposing"), the data in (167) seem to suggest another analysis, involving subject postposing in a manner precisely parallel to SI in French. It is easy to see that such an analysis accounts straightforwardly for why both an aspectual auxiliary and a main verb follow the OI trigger in structures such as (167); just as in SI structures in French, the subject postposes, adjoining to VP. French equivalents to (167) are given below:

(168)a. ?Avec qui pourra Jean aller a New York ?
With who will-be able to-go to

"Who will John be able to go to New York with ?"

b. *Avec qui pourra aller Jean a New York ?

c. ?Qu'est-ce qu'a organizé le peuple ?
What have organized the people

"What have the people organized ?"

d. *Quest-ce que le peuple a organisé ?

There are further striking parallels between SI and OI. Torrego points out that in Spanish certain adverbs may appear in preverbal position in FI contexts but not in OI contexts, as illustrated below:

(169)a. Siempre lee lo mismo Maria.
Always reads the same

"Maria always reads the same thing."

Qué (*siempre) lee Maria ?
What

"What does Mary always read ?"

Now, standard French also (somewhat marginally) permits a preverbal adverb, but never in SI contexts:

(170)a. ?Marie toujours lit la même chose.
always reads the same thing

"Marie always reads the same thing."

b. Qu'est-ce que (*toujours) lit Marie ?
What

"What does Marie always read ?"

Once again, SI and OI seem to pattern together. In view of the observed parallelism between the two rules, we will henceforth treat the latter as a case of subject postposing. As we shall see, this treatment has further desirable consequences for our typology of SAI-type movement rules.

So far, then, we have seen that OI in Spanish is a case of subject post-position triggered by a sentence-initial topic or WH-operator; as such, it clearly falls into the class of SAI-type rules which we have been examining, and more particularly the subclass which includes SI in French. Now, recall that OI fails to show one of the typical asymmetries which we have

associated with predication-related inversion: i.e., the root/non-root asymmetry: in Spanish, as in Catalan, Portugese and Romanian - but unlike in Italian - OI takes place in any tensed clause, whether root or non-root. Moreover, there is no difference in this regard between complements headed by WH-phrases and those which contain topicalization structures. This is illustrated in (171) below:

(171)a. No sabía qué querían esos dos.
NEG know-1s.pst what want-3pl.pst those two

"I didn't know what those two wanted."

b. *No sabía qué esos dos querían.

c. Dinero, sabía qué querían esos dos.
Money

"Money, I knew those two wanted."

d. *Dinero, sabía qué esos dos querían.

The data in (171) would perhaps be surprising if OI were to be analyzed as a V-preposing rule, parallel to SAI and INFL2; however, we have already seen in 4.45 that SI in French shows no root/non-root asymmetry, and once again OI behaves in a precisely parallel fashion, reinforcing our conclusion that both rules are cases of subject postposing.

Of course, this to a certain extent begs the question of <u>why</u> predication-related subject postposing rules should not exhibit the asymmetry which we have seen associated with predication-

related [I,V] preposing rules. In our discussion of SI, we circumvented this question, noting that SI was optional in subordinate clauses and thus irrelevant to the specification of AGR-domains in French. This option is not however open to us with respect to the complete lack of a root/non-root asymmetry shown by OI; it appears that we must extend our parametric analysis of AGR-domains to encompass it. In particular, it appears that both topicalized phrases and WH-phrases must count as (A'-type) P-subjects in embedded clauses in Spanish, which in turn means that neither can be contained in a higher AGR-domain. While this is expected for topicalization structures, which, as we have seen, induce inversion in embedded contexts in a number of languages, it is unexpected for WH-phrases, which we have argued are typically contained in the AGR-domain of a higher clause when contained in an embedded COMP. Interestingly enough, however, there is evidence that fronted WH-phrases in embedded questions in Spanish differ in position from those in the other languages we have been considering. Whereas WH-phrases usually precede complementizers in languages which allow lexically realized "doubly-filled COMP" structures - e.g. in Dutch, Bavarian, Swedish, and Norwegian - in Spanish, they <u>follow</u> the complementizer. This can be seen in structures involving verbs of saying, where a complementizer can be optionally present along with a fronted WH-phrase, as exemplified in (172) below (from Rivero 1980):

(172)a. Te preguntan que para qué quieres el prestamo.
you(OCL) ask-3pl. that for what want-2s. the loan

"They ask you what you want the loan for."

b. *Te preguntan para qué que quieras el prestamo.

In order to account for this data while maintaining the cross-linguistic generalization (first made explicit in Chomsky and Lasnik 1977) that WH-movement takes place to the left of the position occupied by a lexical complementizer, we will assume that Spanish has the option of adjoining WH-phrases to IP. We will also assume that verbs of saying in Spanish may take [-WH, +Q] complements; this fills in the "missing" feature combination alluded to in 4.33 above, and allows us to account for the fact that the complements in question are interpreted as indirect questions. Given the possibility of WH-adjunction to an embedded IP, WH-phrases may occupy the same structural position as embedded topicalized phrases in Spanish, in which case they will not be contained in a higher AGR-domain and will become A'-type P-subjects when fronted at S-structure, inducing OI. (Note that the "S-adjunction" possibility proposed here for WH-phrases has been independently suggested by Lasnik and Saito 1984 in their discussion of languages such as Polish which allow multiple WH-phrases to be fronted in a single clause.)

Interesting supporting evidence for the S-adjunction analysis in Spanish indirect questions is provided by Trentino,

a CI language with FI. Trentino, like standard Italian and unlike Spanish, shows a root/non-root asymmetry with respect to SAI-type inversion; however, unlike standard Italian, Trentino permits a "doubly-filled COMP" under certain circumstances. When a WH-phrase and a complementizer surface together, the former always precedes the latter, as illustrated in (173) below (from Safir 1985):

(173)a. Ho domando cosa che el magna.
Have-1s. asked what that he(SCL) eats

"I asked what he ate."

b. *Ho domando che cosa el magna.

The data in (173) are to be expected if the correlation between non-root inversion and S-adjunction of WH-phrases is genuine.

Thus we see that the absence of a root/non-root asymmetry in OI structures in Spanish can be accounted for by appealing to an independently established parameter, which allows WH-phrases to adjoin to $[I,V]^{max}$ and to become A'-type P-subjects in embedded clauses.

The fact that OI takes place indiscriminately in both root and non-root clauses in fact leads to some interesting additional conclusions. To start with, as Torrego points out, successive cyclic movement triggers OI in each cycle, which can then serve as a diagnostic for the presence of an "intermediate landing site"; this can be seen in (171c.) above, where a

topicalized phrase originating in the most deeply embedded clause forces inversion in both the embedded and matrix clauses. Torrego uses this diagnostic to argue for an analysis of subjacency in Spanish similar to that provided by Rizzi (1982) for Italian, though there is much speaker disagreement as to whether inversion is necessary in every clause where successive cyclic movement has taken place. Of more direct relevance to our investigation, note that OI in this respect resembles both INFL2 in German and SI in French; aside from providing more evidence for the close relation between all three rules, this seems to suggest that our tentative analysis of A'-type predication as applying successive cyclically might well be on the right track.

This concludes our discussion of OI, and, indeed, our cross-linguistic examination of SAI-type rules. I hope to have shown in the preceding sections that the predication conditions introduced at the beginning of this chapter have the explanatory potential to account for a significant range of inversion constructions in several languages. In the next section we will attempt to provide a more explicit parametrization of the cross-linguistic variation we have been examining, before going on in Chapter 5 to examine data from language acquisition.

4.5 Principles and Parameters of SAI-type Inversion

4.50 Introduction

In the following sections, we will summarize the results of our cross-linguistic investigation by proposing an explicit set of parameters governing predication-based inversion rules. First of all, however, let us repeat the central principles which we have been employing in our account of SAI-type rules, beginning with our reformulated statement of P-licensing condition II:

(174) Every AGR-domain must/may have an x-type P-subject.

As indicated in (174), there are at least two types of parametric variation associated with the P-subject Principle: firstly, a P-subject may be optional or obligatory; second, languages may require either an A-type or A'-type P-subject. These variations will be the subject of section 4.51.

Next, we turn to the notion of AGR-linking, which is required by P-licensing Principle III:

(175) Every P-subject must be AGR-linked.

AGR-linking is defined once more as in (176):

(176) AGR-linking: A P-subject α is linked to AGR_i iff α

minimally governs AGR_i.

In the languages we have been studying, neither (175) nor (176) is subject to parametric variation; this may be an artifact of the "division of labour" between the various principles we have proposed in this chapter, or alternatively it might represent the discovery of a genuine universal. We will not be so rash as to proclaim the latter at this preliminary stage of research, though, obviously, "'tis a consummation devoutly to be wished".

Finally, let us return to the notion of AGR-domain first introduced earlier in this chapter:

(177) α is in the domain of AGR_i iff α is in the maximal c-command domain of AGR_i, where AGR_i is an independent AGR, and there is no β, β a barrier for α, such that β excludes AGR_i.

Once again, there is no parametric variation associated with the core notion of AGR-domain, which of course incorporates the fundamental structural concept of government. Note that by incorporating the dependent/independent distinction into the definition of AGR (as in (94) above) we have dispensed with a putative source of variation.

We have identified four other sources of parametric variation in our cross-linguistic survey. The first is what we

might call the "AGR-in-COMP" parameter, which distinguishes languages which appear to observe the No Complementizer Condition from those which do not. The second involves the possibility of WH-adjunction to IP in the syntax, attested in Spanish (and probably in languages such as Rumanian, Polish and Bulgarian which allow multiple WH-phrases to be fronted in a single clause). Third, we have observed extensive variation in what type of element is permitted to act as an A'-type P-subject when fronted at S-structure. Finally, parametric variation also appears to characterize the relationship between matrix/non-matrix clauses and dependent/independent clauses. These four types of parametric variation will be dealt with in sections 4.52, 4.53, 4.54, and 4.55, respectively.

4.51 Parametric Variation and P-licensing condition II

As mentioned above, there are two main types of variation associated with the P-licensing condition II. We will deal with them one at a time.

First of all, let us examine the "optionality" parameter. We have seen that INFL2 languages must be distinguished from the others in our survey on the basis of the following parameter:

(178) In INFL2 languages, every appropriate AGR-domain must contain an <u>A'-type</u> P-subject. In English, French, and

Spanish this requirement need not be met.

Evidence for (178) is provided by the contrasting behaviour of "subjects" in the two sets of languages. As we have seen, even in (Type II) SVO INFL2 languages, surface root-clause SVO order involves the (often string-vacuous) movement of the NP in external argument position to an A'-position; on the other hand, in English, there is evidence (principally from Chung and McCloskey 1983) that this type of vacuous movement need not and does not take place.

We are now in a position to examine what types of evidence might be implicated in the setting of parameter (178). Suppose as an initial hypothesis, we were to adopt the Vacuous Movement Hypothesis (VMH) of Chomsky (1986b.), repeated below:

(179) Vacuous movement is not obligatory at S-structure.

In conjunction with the Last Resort Principle (syntactic rules apply only when forced by grammatical principles), (179) will suffice for English, but obviously not for INFL2 languages. We might expect then that Primary Linguistic Data (henceforth PLD) of a type accessible to first language learners should be available in INFL2 languages to set the marked value of parameter (178) - i.e., to allow vacuous movement. Recalling our rough criteria for acessible PLD - that it constitute

degree-0, type-frequent, positive evidence - we are now in a position to ask whether the relevant type of evidence is available in INFL2 languages.

It turns out that the type of direct evidence needed is only available in clauses of degree-1 embedding in at least some INFL2 languages, thus failing at least one of our three criteria. Type II (SVO) INFL2 languages, as we have seen, have vacuous movement of external arguments, accompanied by fronting of a tensed [I,V] constituent. However, evidence for this is provided only by the different position of negative particles and adverbials in root and non-root clauses in Swedish and Norwegian, and by extraction asymmetries from embedded clauses in Yiddish (see 4.44 above); both of these sources of evidence are of degree-1 embedding. It thus appears that simply treating vacuous movement as the marked value of a binary parameter runs into trouble on learnability grounds.

In fact, there are additional reasons to treat the VMH solution with suspicion. If vacuous movement were simply a marked option, we would expect it to be a marked option in INFL2 languages generally, as well as in English-type languages. But this does not appear to be the case; every INFL2 language we have examined has vacuous movement, whether Type I (S-VP-I), Type II (S-I-VP) or Type III (I-S-VP). Of course, it could always be claimed that this is the result of the historical relationship between INFL2 languages rather than of

some intrinsic property which they share; however, within the context of first language learning, such historical connections should be quite irrelevant, since they are obviously unavailable to the language learner. We conclude that even a parametrized version of the VMH cannot account for the contrast between INFL2 languages and English-type languages with respect to the optionality of A'-type predication.

Given, then, that no appropriate PLD is available to set parameter (178), we are forced to the conclusion that it must be set <u>indirectly</u>, that is, through a set of more accessible PLD which acts as a "trigger" while not necessarily bearing directly on the parameter itself. What might constitute such PLD ?

One possibility is that the unrestricted nature of inversion-inducing elements in the INFL2 languages versus the restricted nature of inversion-inducing operators in a language such as English might provide the appropriate evidence. However, note that topicalized phrases in Spanish also induce inversion, as pointed out in 4.46 above; but Spanish is not an INFL2 language, as shown by sentences such as (180) below:

(180) Vino Juan ayer de tarde.
Came yesterday late

"Juan came late yesterday."

As argued by Jaeggli (1982), following the analysis developed

by Burzio (1981) for Italian, postverbal "subjects" in structures containing unaccusative verbs such as that in (180) are best analyzed as occupying direct object position (i.e., sister to the verb). Now, in that case, either there is an empty pleonastic pro in preverbal subject position (as suggested by the analyses of Chomsky 1982 and Safir 1985) or there is nothing at all in preverbal position (as proposed by Borer 1986). In either case, it cannot be claimed (without the liberal use of unlicensed empty categories) that there is an A'-type P-subject present in structures such as (186). We conclude that Spanish does not have obligatory A'-type predication, in spite of the fact that topicalized phrases automatically trigger OI. This in turn forces us to the conclusion that the latter property is not a diagnostic for an INFL2 language, and cannot serve as PLD for the child attempting to set parameter (178).

Another possible source of evidence is provided by the existence of A'-type pleonastics in INFL2 languages, such as German es and Icelandic thad. However, it must be shown that all INFL2 languages have A'-type pleonastics in order to maintain that they can provide "triggering" data for (178). Unfortunately, this does not appear to be the case: we will now show that at least one INFL2 language - standard Dutch - has no pure A'-type expletive element, in spite of the fact that it is closely related to German and uncontroversially has obligatory

A'-type predication.

As a first step, let us review our arguments for the existence of A'-type pleonastics. First of all, in INFL2 languages which either show (Type I) NCC or (Type II) root/non-root asymmetries, the distribution of pleonastic elements should mirror the relevant predication-based asymmetry. However, care must be taken to separate out "pure" A'-type pleonastic elements from "quasi-argument" pleonastics (see Chomsky 1981), which typically occur with "weather verbs" and show no predication-based distributional asymmetries in any of the languages under consideration. Pure A'-type pleonastics, on the other hand, are only required by the P-subject Principle; it follows that in a Type I language such as German, a pure A'-type pleonastic should not occur together with a lexical complementizer, since, as argued in 4.44 above, the latter contains dependent AGR which is linked to a matrix AGR-domain, and therefore does not require a P-subject. As argued by Safir (1985), this prediction is borne out for at least some German speakers, whose judgements are illustrated in (181) below:

(181)a. (Es)* wird getanzt werden.
It will danced to-be

"There will be dancing."

b. Er sagte, daβ (es*) getanzt werden wird.
He said, that

"He said there will be dancing."

The second major diagnostic for an A'-type pleonastic in INFL2 languages is its failure to occur in questions, including yes/no questions; both types already contain a P-subject and thus satisfy P-licensing condition II without recourse to an A'-type pleonastic. Once again, German es conforms to this pattern, as we have already seen in 4.43.

Now, it can be shown that Dutch, which otherwise closely resembles German, has no pure A'-type pleonastic. Dutch has two pleonastic elements, het and er, roughly equivalent to English it and there, respectively (see Travis 1984 for an account of the differences between these two types, though under rather different theoretical assumptions). Standard Dutch does not allow either to be missing in embedded contexts, as illustrated in (182) below:

(188)a. (Er)* wordt gedanst.
There was danced

"There was dancing."

b. Hij zei dat (er)* wordt gedanst.
He said that

c. Het is duidelijk, dat Johan ziek is.
It is clear that Johan sick is

"It's clear that Johan's sick."

d. Hij zei dat (het)* duidelijk is, dat Johan ziek is.

Moreover, neither er nor het disappears in questions, unlike

es:

(183)a. (Waar) wordt (er)* gedanst ?
(Where) was there danced

"(Where) was there dancing ?"

b. (Warum) is (het)* duidelijk, dat Johan ziek is ?
(Why) is it clear that sick is

"(Why) is it clear that Johan is sick ?"

We must conclude that the behaviour of pleonastic elements in INFL2 languages cannot serve as triggering evidence for obligatory A'-type predication, since pleonastics in Dutch, as opposed to those in German, do not show distinctive A'-type behaviour. (Note that this has no direct bearing on the question of optional versus obligatory A'-type predication. If anything, the contrast between the Dutch and German data shows, as proposed by Safir (1985) that Dutch has obligatory assignment of nominative Case, whereas German does not; both have obligatory A'-type predication.)

So far, then, we have shown that neither of the two most salient properties of INFL2 languages can serve as triggering data for parameter (178). This leaves us in a somewhat paradoxical position, since there appears to exist neither direct nor indirect triggering evidence for obligatory A'-type predication.

In order to resolve this paradox, I would like to take a

rather different approach. We have assumed up till now that the INFL2/obligatory A'-type predication phenomenon is marked, a reasonable position given its typological scarcity - see Haider (1986b.) for comments. However, suppose we were to relativize this assumption, as in (184):

(184) If a language has A'-type predication, then, in the unmarked case, it will have obligatory A'-type predication.

Given (184), a child need only learn that A'-type predication is <u>possible</u> in a given language to infer that it is obligatory; in other words, INFL2 languages will turn out to be (relatively) <u>unmarked</u>. On the other hand, languages such as English, French, and Spanish will turn out to be (relatively) marked; the child will need further evidence (in the form of sentences where INFL2 has <u>not</u> applied) in order to learn that A'-type predication is "optional". Such evidence is readily available in English; any of the following types of sentence (all of which meet the PLD-accessibility criteria) will serve to set the marked value for (184):

(185)a. Yesterday, I went to the movies.

b. Dogs I like, but people are another matter.

c. How come I can't eat worms ?

d. Who cares ?

(185a.) and (b.) contain topicalized elements which fail to trigger INFL2; (185c.) an exceptional non-SAI inducing WH-phrase; and (185d.) a subject question. The last of these examples is particularly interesting from the point of view of learnability, since it involves a type of indirect evidence which nevertheless conforms to conventional "positive-evidence-only" restrictions: since in English only an auxiliary verb may front to pre-sentential position, it is the lack of auxiliary do in sentences such as (185d.) which alerts the language learner that SAI has not taken place, and therefore that the subject WH-phrase has itself not been fronted to an S-external A'-position; this in turn tells the learner that an A'-type subject cannot be obligatory in English.

Turning to French (and other CI languages), the marked value of (184) is learnable from examples parallel to (185a.) and (185d.), as well as questions without any inversion at all, which are quite frequent in Colloquial French, especially with SCLs and other "light" subjects:

(186)a. Hier soir, je suis allé au cinéma.
Yesterday evening I(SCL) am gone to-the cinema

"Yesterday evening I went to the movies."

b. Qui est allé au cinéma ?
   Who is

   "Who went to the movies ?"

c. Où tu vas ?
   Where you(SCL) go

   "Where are you going ?"

(Note that (186b.) provides the same type of indirect evidence as subject questions in English: the absence of a SCL, generally obligatory in CI, alerts the language-learner that the WH-phrase is in situ.)

Finally, in OI/FI languages such as Spanish, "unaccusative" FI structures such as that in (180) will serve as evidence for the optionality of A'-type predication; since the [I,V] constituent in such structures must be generated inside "VP", AGR features must also be inside VP, and the "inverted" NP must be in an A-position; the learner can conclude that A'-type predication is not applicable. Questions containing non-OI-inducing WH-phrases such as those in (187) will also serve as evidence for the optionality of A'-type predication:

(187)a. Por qué Juan quiere salir ?
   Why wants to-leave

   "Why does Juan want to leave ?"

b. Cuando Juan dio el libro a Maria ?
   When gave the book to

   "When did John give the book to Mary ?"

Thus we see that when reformulated as in (184), the "optionality" parameter (178) meets appropriate learnability requirements.

Let us then move onto the second putative source of cross-linguistic variation associated with the P-subject principle: A-type versus A'-type predication.

Whereas it is clear even from our limited investigation that A'-type predication is somewhat marked (most languages, after all, do not have an INFL2 condition nor operator-induced inversion) it may well be that A-type predication is universal, or at least highly unmarked. It would take us far too far afield to attempt to either confirm or repudiate this claim, since our interest has been focussed on A'-type and not A-type predication; however, we conjecture that at least in languages with AGR features, A-type predication is obligatory. In this regard, we repeat the observation made earlier to the effect that morphological A-type subject-verb agreement is very frequent cross-linguistically (see Roberge 1986 for some typological remarks), whereas morphological A'-type "topic-verb" agreement is either vanishingly rare or altogether absent (see Li and Thompson 1976); though this is probably partly a reflection of the different grammatical levels at which the two types of predication apply, it also surely reveals a built-in markedness asymmetry.

Note that the structural conditions on A-type predication

cannot be identical to those on A'-type predication. This is because for A-type predication, the notion of P-subject is determined by thematic considerations, at least in "subject-prominent" languages such as those we have been investigating. Essentially, the NP which controls AGR-features in such languages is the "most external" argument on a thematic heirarchy - or sometimes, when an argument with a certain θ-role need not or cannot be "externalized", an expletive argument with no θ-role. We will assume here that FI languages such as Italian and Spanish differ from those with Indefinite Inversion such as English and French in allowing an NP with an internal (theme) argument to act as the A-type P-subject for AGR-features generated in [I,V]; this is partly accounted for by the different positions occupied by [I,V] in FI and II languages - recall that the former may generate [I,V] in VP, the latter may not and therefore have V-to-I raising. Obviously, this is merely a sketch of A-type predication; much further work needs to be done, which we cannot undertake in this already over-extended study.

Turning to the question of language-learning, it seems fairly clear that evidence for A'-type predication is readily available in languages which do have it: in English, such evidence is provided by the operation of SAI, in INFL2 languages by [I,V] movement, in French by CI and SI, and in Spanish by OI. Whether - and at what stage in acquisition -

children actually do make use of this evidence is another matter, as we shall see in section 4.6 below.

4.52 The AGR-in-COMP parameter

In section 4.44, we introduced the concept of "dependent" AGR in COMP, which characterizes those INFL2 languages which appear to obey Goldsmith's (1981) No Complementizer Condition (NCC). When specified for AGR features, complementizers in such languages count as dependent AGR, by condition (98), repeated below as (188):

(188) AGR is dependent if and only if it is contained in another AGR-domain.

In languages where complementizers cannot be specified for AGR features, each instantiation of AGR in INFL will be independent, since even in embedded clauses INFL will not be in the government domain of a higher AGR; CP (=S') will count as a barrier to government in the sense of Chomsky (1986b.). On the other hand, AGR-in-COMP will be in in the government domain of a higher AGR, and AGR-in-INFL in a clause introduced by AGR-in-COMP will in turn be in the government domain of AGR-in-COMP; both will be dependent, by (188). In this way, we account for the NCC effect in languages which allow AGR-in-COMP.

Now, the question arises as to whether AGR-in-COMP is a freely chosen option amongst the languages we have been examining. In our cross-linguistic survey, it seems confined to Type I (S-VP-INFL) INFL2 languages. If the AGR-in-COMP parameter were quite independent, we might expect languages to exist with only optional inversion in complementizer-less clauses, no inversion in clauses with lexical complementizers, and possibly morphological reflexes of AGR-in-COMP. Though CI in Standard French is compatible with these specifications, it is equally compatible with an analysis parallel to that of English SAI; the lack of complementizer morphology and the fact that complementizers cannot be deleted in French deprives us of any crucial deciding evidence.

Before leaving the question, however, let us turn to the question of what kind of evidence might be required to set the AGR-in-COMP parameter.

Obviously, in languages such as Flemish or the Bavarian dialect of German, overt morphology on the complementizer counts as positive evidence for AGR-in-COMP. Though this evidence must be at least of degree-1 embedding, since complementizers are not normally present in matrix clauses, notice that it is <u>lexically-based</u>; that is, it is tied to lexical properties of the complementizer system, which must in any case be learnt by the child. Where no complementizer morphology is visible, the operation of INFL2 in embedded

complementizer-less clauses will also provide positive evidence; once again, this evidence will be of degree-1, but again, tied to properties of the complementizer system (specifically, to the "deletability" of complementizers). This leaves languages in which neither type of evidence is available: Dutch and Frisian are of this type, as we have seen. Dutch contains a set of "weak" clitic pronouns which might provide evidence for AGR-in-COMP; however, notice that French SCLs also attach to complementizers, yet we have not been treating French as an NCC language. We have also seen that Frisian has partly lost its AGR-in-COMP properties, accounting for "exceptional" inversion structures in that language. It appears, then, that direct positive evidence of the required type might be missing in at least one AGR-in-COMP language, leading to an "unstable" NCC system. Of course, the fact that inversion in embedded clauses is <u>exceptional</u> in Frisian (as documented in deHaan and Weerman 1986 and in 4.44 above) indicates that Frisian is still a Type I (NCC) language; and embedded inversion in clauses containing complementizers of course itself counts as positive evidence just in case the language learner "expects" an NCC-type system. This in turn indicates that the NCC is <u>unexceptional</u> in the context of first language learning for Type I languages; which in turn leads us to the conclusion that some independent property of Type I languages leads the child to expect NCC-type properties; in

other words, AGR-in-COMP is a dependent and not an independent parameter.

What might the required property be ? The most obvious answer is that the S-VP-INFL order of Type I languages is tied to the cluster of properties we have associated with the NCC. Counter-evidence to this claim would of course be provided by an NCC language with different major-constituent order; French is a putative counter-example, but provides no deciding evidence.

4.53 The WH-adjunction parameter

We have distinguished OI in Spanish (and Portugese and Rumanian) from the SAI-type processes we have been investigating in other languages on the basis of the ability of WH-phrases to adjoin to IP (= S) in embedded clauses. Supporting evidence for this conjecture comes from the appearance of WH-phrases in the complements of certain Spanish verbs in a position directly following a lexical complementizer, a position shared by topicalized phrases. We further speculated that the verbs which allowed this type of complement were specified [-WH] (optionally) but [+Q] (recalling our discussion of topicalization and exclamatives in English); this account has the advantage that (a) it explains a principled class of "Doubly filled COMP filter" violations in

Spanish; (b) it accounts for a set of exceptions to Chomsky and Lasnik's (1977) claim that WH-movement is universally to the left of the lexical complementizer position; (c) it fills in a gap in the paradigm created by the cross-classification of the features [±WH] and [±Q]; (d) it accounts for OI with WH-phrases in embedded contexts, since when adjoined to IP, WH-phrases, like topicalized constituents, become A'-type P-subjects and are therefore expected to induce inversion.

It should be pointed out that adjoined WH-phrases and topicalized phrases cannot be completely identified, however. Rivero (1978) shows that whereas in Spanish both WH-movement and topicalization are successive cyclic processes which in general obey subjacency, only WH-movement obeys the WH-island condition; topicalization, on the other hand, obeys a "topic-island condition", indicating that there is a more general "like-constituent island constraint" at work, and more to the point, that WH-phrases cannot be identified with [-WH] topicalized constituents.

Now, obviously not every language allows WH-adjunction to IP in embedded clauses. We have observed that, for the most part, embedded topics and embedded WH-phrases show systematically different behaviour with respect to SAI-type inversion; this is true in all the Germanic languages except possibly for Icelandic, and - in a rather different fashion - for English as well. In fact, it is even true for Spanish,

since the WH-complementizer si, as well as "non-argument" WH-phrases, fail to trigger OI. Let us accordingly assume, as a first hypothesis, that WH-adjunction is a marked option. If so, we need - or rather, the first language learner needs - accessible positive evidence to learn the WH-adjunction parameter in a language such as Spanish. Now it might be thought that such evidence is provided very directly by the application of OI in embedded clauses. Recall, however, that the only way to tell whether OI applies is by either direct or indirect negative evidence; clauses in which OI has not applied are marked as ungrammatical. Moreover, such evidence, once again, is only available in embedded clauses; it thus appears that the child learning Spanish must apply degree-1 negative evidence to the task of learning OI - an unpleasant conclusion.

The obvious move to make in such a situation is to ask whether there is another source of accessible evidence for the WH-adjunction parameter. One possibility is that overt (complementizer + WH-phrase) sequences are the triggering evidence. While such sequences are clearly confined to embedded clauses, and thus still count as degree-1 evidence, it is at least positive evidence. As for our third accessibility criterion, that of frequency, it is difficult to ascertain the relative abundance of (complementizer + WH-phrase) sequences in colloquial spoken Spanish, though they certainly occur in the complements of relatively token-abundant verbs. Note that if

(complementizer + WH-phrase) sequences are crucial triggering evidence, we might expect to find them in Rumanian, Portugese, and Catalan as well; unfortunately, I am not currently in a position to investigate this prediction further.

4.54 Possible P-subjects

So far, we have investigated languages in which P-subjects are unrestricted (INFL2 languages) and those in which only a lexically restricted subclass of fronted elements may be P-subjects (OI languages, CI languages, SAI languages). It is clear that in the latter class (which we have labelled "optional" A'-type predication languages) the range of elements which trigger inversion (i.e., act as P-subjects) does not arbitrarily vary from language to language; it seems that, given A'-type predication, there are certain sets of closed class items which are most likely to be marked "[+P]". The following table summarizes the range of inversion-triggering elements in some of the languages we have been investigating:

(189)

	INFL2	Spanish	French	English
XP	yes	yes	-	no
WH(I)	yes	yes*	(yes%)	yes%
WH(E)	yes	yes	no	no
PO	yes	yes	yes%	yes

(WH(I) = interrogative WH-phrase, WH(E) = exclamative WH-phrase, PO = Propositional operator. (*) indicates that only

argument WH-phrases trigger inversion, (%) that there are lexical exceptions, and a blank space that fronting is inapplicable in a given context.)

Note first of all that, as predicted by the Designation Convention - see (74) in section 4.33 above - "exceptional" behaviour is confined to a certain class of closed-class items - i.e., "operators" in a general sense that includes WH-phrases and certain propositional operators, as discussed in section 4.23. From the point of view of language acquisition, the implications of this are obvious; as soon as the child finds a fronted "non-operator" XP triggering inversion, then s/he will know that every non-operator will trigger inversion. However, as can be seen from the behaviour of non-argument WH-phrases in Spanish, "XP-induced" inversion does not guarantee uniform application of "operator-induced" inversion; the latter may be lexically restricted in a language where the former applies.

Let us then focus on operator-induced inversion, beginning with "WH-operators". As (189) shows, English is by no means exceptional in the exceptional behaviour of its WH-operators. The obvious question to ask with respect to such exceptionality is whether it falls into regular sub-patterns, or whether it must be simply treated as "purely" exceptional. In terms of learnability, this question can be cast in a slightly different way: does the child proceed lexical item by item (thus avoiding overgeneralization and subsequent problems of retreat) or

alternatively, does the child employ semantic or syntactic subregularities (in which case, we might expect limited overgeneralization) ?

Exactly the same questions apply to inversion-inducing propositional operators. Given, for example, that perhaps fails to trigger SAI in English, yet its French equivalent peut-être triggers CI, we might expect data from language acquisition to to throw some light on the questions of whether the child learns A'-type P-subjects lexical-item-by-lexical-item or subclass-by-subclass, and of what might constitute the basis of the relevant subclasses. We will return to these questions in some detail in section 4.6, where we will examine data from real-time language acquisition.

4.55 The Dependent/Independent Clause Parameter

Our final parameter is perhaps the least well-defined. In our discussion of colloquial French matrix questions with overt complementizers (see section 4.45), we suggested that such structures were a marked class of "dependent matrix" clauses, and further speculated that we might find cross-linguistic variation as to how closely the matrix/subordinate distinction correlated with the dependent/independent distinction. Logically, in fact, there are sixteen possibilities, though most are ruled out by a priori considerations, as shown below:

(190)

	IM	IS	DS	DM
a.	+	+	+	+
b.	+	+	+	-
c.*	+	+	-	-
d.*	+	-	-	-
e.*	-	-	-	-
f.	+	-	+	+
g.*	+	-	-	+
h.*	-	-	-	+
i.	+	-	+	-
j.*	+	+	-	+
k.*	-	+	-	+
l.*	-	+	-	-
m.*	-	+	+	-
n.*	-	-	+	+
o.*	-	-	+	-
p.*	-	+	+	+

IM = Independent Matrix, IS = Independent Subordinate, DS = Dependent Subordinate, DM = Dependent Matrix. (*) = an impossible combination.

As (190) shows, only four out of the sixteen logical possibilities represent possible natural-language systems. We have excluded the others on the basis of the following considerations:

(a) There is a conditional relation between IM and DM, and between DS and IS; in each case, the former implies the latter, but not vice-versa. This excludes possibilities (c., h., j., k., l., n., p.).

(b) No language lacks matrix clauses: this excludes (e.), (m.), and (o.), in addition.

(c) No language lacks subordinate clauses: this excludes (d.). and (g.) in addition.

This leaves possibilities (a.), (b.), (f.), and (i.). The next question is, are there actual languages corresponding to such putatively possible systems ? Before dealing with each possibility in turn, however, let us first set up criteria for distinguishing dependent and independent clauses.

The first and most obvious of these criteria is the existence of lexical complementizers in dependent but not independent clauses. Complementizers, as we have argued elsewhere in this thesis, are quintessentially subordinating elements which are subcategorized by governing predicates. Even in languages which permit dependent matrix clauses, complementizers seem to be confined to non-declarative environments (typically questions, as in French and Igbo); one interpretation of this distribution is that non-declarative clauses typically involve an operator-like element which can act as a "pseudo-predicate", and thus subcategorize a subordinating complementizer.

The second criterion is the failure of rules which are confined to independent clauses to apply in dependent clauses. Notable amongst such rules are those we have been investigating: SAI, INFL2, CI. In NCC languages, the failure of

inversion rules will always correlate with the existence of either a complementizer or a WH-phrase, or both, in matrix clauses; however, this is not necessarily the case in languages which show a root/non-root asymmetry, where dependent clauses may or may not have a complementizer.

Conversely, independent clauses in non-matrix environments will never be headed by complementizers and will typically undergo "root" rules. Let us then return to the question of which possible systems are instantiated cross-linguistically.

System (a.) treats the matrix/subordinate and dependent/independent parameters as entirely separate. In a language employing (a.), we will expect to find two types of matrix clause, dependent and independent, and two types of subordinate clause, dependent and independent. In system (b.), dependent clauses may occur in matrix positions, but independent clauses may not occur in subordinate positions; system (f.) shows a converse pattern, where both dependent and independent clauses occur in subordinate environments, but only independent clauses occupy matrix positions. Finally, in system (i.), the two parameters are indistinguishable: a dependent clause is equivalent to a subordinate clause, and an independent clause to a matrix clause.

Now, let us re-examine some the languages we have been investigating in order to determine which of the four patterns detailed above they might correspond to:

(191)

	IM	IS	DS	DM
Standard English	+	-	+	-
Colloquial English	+	+	+	-
Standard French	+	-	+	-
Colloquial French	+	-	+	+

As shown by (191), three out of the four possible systems are instantiated by dialects of French and English alone. A possible example of the "missing" system is provided by the Tsimshian language family of N.W. B.C.. In the study of Tsimshian languages, there is a well-established distinction between dependent and independent clauses, both of which may occur in matrix or subordinate environments; dependent clauses are typically headed by a propositional operator (generally NEG or a modal element) and are distinguished from independent clauses by a number of syntactic tests: see Hunt (forthcoming) for details.

Assuming that all four possible systems are instantiated, the next question that faces us is one of markedness. Clearly, IM and DS clauses are unmarked; however, it is not so obvious that IS and DM clauses are marked, though this does seem the most plausible possibility. Learnability considerations in fact lead to the same conclusion; if DM clauses were unmarked, then only negative evidence would tell the language-learner that sentences such as those in (192) were ungrammatical:

(192)a. *What you did ?

b. *If you went ?

c. *Why to stay ?

In fact, as we shall see, sentences such as (192a.) are regularly produced by children at a certain stage in the acquisition of English; however, the other two types of matrix dependent clauses are not, leading us to suspect that cases of (a.) in language acquisition are not "mis-settings" of the dependent/independent clause parameter. Note in addition that even in standard English, where dependent questions are not as a rule allowed to occupy matrix positions, the following dependent clauses are at least semi-grammatical in matrix environments:

(193)a. Oh, to be in England, now that April's here !

b. Not that I'll return until the ice-queen leaves.

c. If only the spell were broken !

(193) shows us that the difference between Standard English and, say, Colloquial French, which allows a wider range of DM clauses, is not absolute, but a question of degree; to this extent, the table in (191) above is a categorical idealization of a continuum, with one end occupied by a language which allows very few or no dependent clauses in matrix positions,

the other by a language which allows a wide range or possibly all dependent clauses to surface in matrix environments.

Exactly the same point can be made about IS clauses. We have seen that in general SAI in subordinate clauses is ungrammatical, but we have also seen that colloquial English allows such IS structures, which are particularly widespread in NNE dialects, as attested by the work of Labov et al. (1968), discussed in 4.33 above. Once again, it appears that SE and NNE represent two points on a continuum rather than being differentiated by a binary parameter.

Let us look more closely at this notion of a "parametric continuum". There are in fact two possible ways of looking at such a notion. The first, more literal interpretation predicts that languages are arranged along an implicational continuum; according to this view, a given language will only contain a DM or IS structure of markedness value (n) if that language also contains all DM/IS structures of markedness value (< n). The second, looser interpretation makes no such prediction; instead, each clausal type has an intrinsic markedness value for DM or IS structures, allowing for the possibility that a language may have "scattered" values for the parameter, with some DM/IS structures set at the unmarked value, others set at the marked value. We will call these the "implicational" and "non-implicational" versions of the parametric continuum.

Now, given that both versions make explicit predictions, we should be in a position to ascertain which is appropriate for the dependent/independent parameter. In particular, counter-evidence to the implicational version would be provided by languages which allowed disjoint sets of clauses in DM/IS structures. The two tables in (194) below shows which IS and DM structures are permitted in some of the languages we have been investigating (lack of data accounts for why we have not included the others):

(194)a. IS structures:

	D	I	E
Standard English	-	-	?
Colloquial English	-	+	?
Standard French	-	-	?
Colloquial French	-	-	?

b. DM structures:

	D	I	E
Standard English	-	-	+
Colloquial English	-	+	?
Standard French	-	-	+
Colloquial French	-	+	+

(D = declarative, I = interrogative, E = exclamative)

(194) is, unfortunately, inconclusive, but certainly suggestive; it seems to show that both DM and IS structures are most likely to occur with exclamatives and least likely to occur with declaratives. We tentatively conclude in favour of

the "implicational" version of the parametric continuum, thouh obviously more work needs to be done here.

The introduction of "continuous" (versus "discrete") parameters such as the dependent/independent parameter has a number of consequences for learnability theory. In particular, it appears no longer adequate to claim that one end of the continuum created by a continuous parameter is marked and the other unmarked; it might be, for example, that the "unmarked" point on the continuum is somewhere in the middle, in which case certain types of dependent clause - such as those in (193) - would be relatively unmarked in matrix positions with a given interpretation, others - such as declarative <u>that</u> clauses, which, as shown in (194), never appear in matrix environments - relatively marked.

Now, all this has obvious implications for learnability theory. Suppose a dependent clause of Type X was relatively unmarked - by UG - in matrix position, yet in a language L it was treated as ungrammatical in matrix environments, In that case, only negative evidence would suffice to learn L.

The most obvious way round this problem would be to claim that language L is unlearnable in the relevant respects; in other words, learnability considerations would effectively cut off the continuum at the point of least markedness. In that case we would expect to find <u>no</u> language with an absolute constraint against DM clauses, or conversely, against IS

clauses. This makes clear predictions both cross-linguistically and in terms of real-time language acquisition: we will review the former very briefly and turn to the latter in Chapter 5.

Within the limited sample of languages we have been investigating, it appears to be correct that no language has an absolute constraint against DM clauses. Haider (1986b.) notes that in German, clauses such as those in (197) are acceptable in matrix environments "...in contexts of exclamation or musing questions":

(197)a. Daβ mir das nicht früher aufgefallen ist !
That to-me that not earlier struck is

"That I did not notice it earlier !"

b. Was du nicht sagst !
What you not say

"What you didn't say !"

Exactly the same types of structure appear to be grammatical in other INFL2 languages; see Weerman (1983) for Dutch. We have already seen that English and French both allow certain similar types of dependent clauses in matrix environments; though I lack relevant data on OI languages, it is my firm prediction that they will turn out to follow the same pattern.

Turning to the question of IS clauses, data is much less clear. Setting aside NCC languages, which allow IS-type

clauses for reasons which have nothing to do with the dependent/independent clause parameter, NNE and Icelandic both alow SAI in embedded clauses with WH-phrases; however, this appears to be a marked property, shared by no other INFL2 language. It may well be that IS clauses are generally more marked than MD clauses, in which case the language learner will start off "further down the continuum" towards an absolute prohibition against IS structures.

4.56 Summary

Before (at last) turning to data from first language acquisition, let us review the principles and parameters which we have proposed to account for predication-linked inversion processes. In effect, we have now constructed a "logical" model of language acquisition, which will allow us to make an explicit comparison between the "expected" course of first language learning and what is known about real-time language acquisition - obviously, once we have abstracted away from performance factors. Let us then very briefly summarize the relevant principles and parameters, beginning with the two most important components of the theory developed here, namely, P-licensing conditions II and III:

(196) P-licensing conditions:

II. Every AGR-domain may/must have an A/A'-type P-subject.

III. Every P-subject must be AGR-linked.

These two principles are supplemented by two crucial definitions: firstly, that of AGR-linking, as in (198):

(197) AGR-linking:

A P-subject α is linked to AGR_i iff α minimally governs AGR_i

and that of AGR-domain, as in (198):

(198) AGR-domains:

α is in the domain of AGR_i iff α is in the maximal c-command domain of AGR_i, where AGR_i is an independent AGR, and there is no β, β a barrier for α, such that β excludes AGR_i.

Of these four central notions, only (196.II) is subject to parametric variation, as far as can be gathered from our (admittedly limited) cross-linguistic investigation. Accordingly, we will relativize (196.II) in the following

manner:

(199) The parametrized P-licensing condition II:

I. Every AGR-domain must have an A-type P-subject.

II. (a) No AGR-domain may have an A'-type P-subject, or
(b) Every AGR-domain must have an A'-type P-subject or
(c) Every AGR-domain may have an A'-type P-subject.

As argued above, (I) is a putative universal; it corresponds to the "extended" part of the Projection Principle, discussed in Chomsky (1982), Rothstein (1983), and elsewhere. On the other hand, (II) contains three options, ordered from least to most marked; as argued in section 4.51, if A'-type predication applies, then in the unmarked case it will apply uniformly, as in INFL2 languages, rather than in a lexically restricted set of cases, as in SAI and CI languages.

We have identified four other putative parameters which interact with the central principles we have proposed. The first is the AGR-in-COMP parameter, which is tied to the distinction between dependent and independent AGR mentioned in (198). We have speculated that AGR-in-COMP might be confined to Type I INFL2 languages; though evidence is hardly conclusive, we will assume here that in such languages, AGR-in-COMP is

unmarked. Accordingly, we will formulate the following parameter:

(200) AGR-in-COMP:

In a Type II (S-INFL-VP) language,
(a) lexical complementizers contain AGR features (unmarked), or
(b) lexical complementizers do not contain AGR features.

Our next parameter is what we have called the "WH-adjunction parameter". Strictly speaking, this is a misnomer; we have claimed that independent clauses lack complementizers, and therefore that WH-movement to the periphery of such clauses is always a case of adjunction. More precisely, then, we might refer to the "dependent WH-adjunction parameter". In 4.53 above we suggested that in Spanish this parameter was related to the ability of a certain class of verbs (verbs of "saying") to select for [-WH, +Q] complements, a possibility not instantiated in English (see 4.34 above); accordingly, we might formulate the parameter as follows:

(201) The dependent WH-adjunction parameter:

WH-phrases may only adjoin to IP in the complements

of verbs which select for [-WH,+Q] complements.

Assuming that the latter property is marked, the former will also be so.

Next we return to the question of "possible A'-type P-subject". Here, it is clear that we are not dealing with a "parameter" in the usual sense of a binary choice-point. Rather, what is at issue is (a) what type of lexical class may function to trigger predication-related inversion, and (b) what type of learning (lexical, categorial, or a mixture) is associated with each lexical class.

We have claimed that an initial delimitation of the notion of possible A'-type P-subject is provided by the Designation Convention of Emonds (1985), which systematically distinguishes open-class from closed-class lexical items (see Chapter 3 and 4.33 above). Given that this distinction is available to the language-learner, the Designation Convention immediately accounts for how (non-operator) XP-induced inversion is learnt; the learner need only encounter one instance of such inversion to conclude that it applies uniformly.

As for operator-induced inversion, the Designation Convention makes no predictions. Clearly, UG must specify that only certain subsets of closed-class items are potential A'-type P-subjects (e.g., WH-question operators, but not WH-relative operators; monotone decreasing operators, but not

monotone increasing operators) since the association of such subsets with SAI-type inversion rules appears to have cross-linguistic validity. It is also clear that, for the purposes of A'-type subject-hood, the subsets are to a degree independent of each other: for example, Spanish exclamative WH-phrases induce OI, but those in French and English fail to trigger CI and SAI, respectively; on the other hand, Spanish adjunct WH-phrases do not trigger inversion, whereas (most of) those in English and French do. At this stage, then, we will limit ourselves to the following unordered list of possible A'-type P-subjects (we have used the feature [±PO] to distinguish inversion-inducing propositional operators from other XPs - no particular theoretical significance is attached to our use of such a feature):

(202) Possible A'-type P-subjects:

(a) [-WH, -Q, -PO]
(b) [+WH, +Q, -PO]
(c) [+WH, -Q, -PO]
(d) [-WH, -Q, +PO]

(a) is the case of XP-induced inversion, (b) that of interrogative-induced inversion, (c) that of exclamative-induced inversion, and (d) that of propositional operator-induced inversion. The last three of these subsets, being closed-class sets (in Emonds' terms, specified for purely <u>syntactic</u> features) may be further lexically restricted. It is possible, and certainly compatible with our cross-

linguistic data, that sets (b) and (c), which are in a subset relation with each other, are implicationally linked; in that case, we might formulate the following statement, where [+P] = "possible A'-type P-subject":

(203) If some member of the set with the syntactic features [+WH, -Q] is [+P], then some member of the set [+WH, +Q] is also [+P].

Finally, let us turn to our last parameter, the dependent/independent clause parameter. Here, we have identified four "possible systems" - but at the same time, pointed out that all of these systems are idealizations and none of them might actually be instantiated in natural languages. The reason for this apparently paradoxical conclusion is that the parameter in question is what we have called "continuous" rather than "discrete" (one is tempted to call it indisreet instead, but discretion is the better part of valour); rather than applying in a uniform manner to every clausal type, it may be set (in the unmarked case) one way for for one type and the opposite way for another type. Thus, as shown in (194), it appears to be (relatively) unmarked for dependent exclamatives to appear in matrix environments, but not dependent declaratives. Formalizing our (admittedly tentative) conclusions with respect to the

dependent/independent parameter, we come up with the following implicational statements, where DM = dependent matrix, IS = independent subordinate, d = declarative, i = interrogative, and e = exclamative:

(204)a. If DM/d, then DM/i.
b. If DM/i, then DM/e.
c. If IS/d, then IS/i.
d. If IS/i, then IS/e.

The implicational statements for DM clauses - (a) and (b) - are identical to those for IS clauses - (c) and (d) - though they are set independently within a given language. This suggests that we might collapse the two, as in (205):

(205)a. If X/d, then X/i;
b. If X/i, then X/e,

where X is a variable over "non-canonical clausal type".

This completes our overview of the parametric apparatus we have set up on the basis of our cross-linguistic investigation of predication-based inversion rules. Now, at long last, we turn to data from first language acquisition.

CHAPTER 5

THE ACQUISITION OF SAI-TYPE INVERSION

5.0 Introduction

In this chapter, we will attempt an explicit application of the parametric model presented in the last section to data from first language acquisition. The data itself will be drawn from a variety of published and unpublished sources; nearly all of it, unfortunately, is from the acquisition of English, though we will refer to several studies on German and one on Dutch.

Inevitably, while there is a considerable quantity of available data on the acquisition of SAI, much of it is inconclusive; as discussed in Chapter 1, the problem of extrapolating from limited corpora to underlying competence can only be resolved by a "multi-methodological single-subject" approach to language acquisition, an approach which has not yet been widely taken. In section 5.33 we will present findings from a pilot study using this type of approach, which, while limited in scope and suffering from a number of practical and theoretical defects, seems to me to present a promising line of inquiry for future work.

The chapter will take the following form: in 5.1, we will make a set of predictions, based on the position of English SAI within the parametrized model established in the preceding

sections. This will serve as the basis for an investigation of the relation between the "logical" learner of English and the "real-time" learner. In 5.2, we will give an overview of the acquisition of "WH-movement", following on from our account of the acquisition of auxiliary verbs at the end of Chapter 3. In 5.3, we will focus our attention on a widespread - and famous - set of acquisition errors involving the interaction of WH-movement and SAI in English, beginning with longtitudinal studies, and turning in 5.4 to cross-sectional studies. In 5.5, we will turn to studies of the acquisition of INFL2 structures in German and Dutch. In 5.6, we will review previous attempts to explain the developmental patterns characteristic of the acquisition of SAI, before providing our own account in 5.7. We will conclude in 5.8 by pointing out some of the discrepancies between the predictions made in 5.1 and the course of real-time language acquisition, and propose an alternative model of syntactic acquisition which I hope will overcome some of the deficiencies of previous accounts. Without further ado, then, let us proceed.

5.1 What the Parameters Predict

5.10 Introduction

In this section, we will attempt to situate SAI in English within a context of cross-linguistic markedness. Ceteris paribus, this should allow us to make specific predictions

concerning the acquisition of SAI, which we will then be able to "test out" against acquisition data. We have identified five main sources of parametric variation in our investigation of SAI-type inversion: let us then examine them in turn.

5.11 P-licensing parameters

The first, source of parametric variation, repeated below as (1), concerns the status of A'-type predication, which may either be precluded, obligatory, or optional:

(1) (a) No AGR-domain may have an A'-type P-subject, or
(b) Every AGR-domain must have an A'-type P-subject or
(c) Every AGR-domain may have an A'-type P-subject.

Values (a. - c.) are in order of markedness, as discussed in 4.51 above. Incidentally, it might appear that the markedness relationship between (b.) and (c.) violates the Subset Principle of Berwick (1982) - see Chapter 1 for details - since at first sight the language (i.e., the set of strings) generated by (c.) forms a subset of the language generated by (b.). However, note that in fact the languages generated by grammars containing (b.) and (c.) are partially disjoint; (c.) will generate instances of fronted constituents without accompanying inversion, for example, whereas (b.) will not. The Subset Principle is in other words irrelevant to (1).

Now, as we have seen, English chooses option (c.), the most

marked parameter of the P-subject principle. In a "linear parametric" model - again, see Chapter 1 for details - this should lead us to expect three stages in first language acquisition: at first, no predication-related inversion rules will apply at all, as the child chooses parameter (a.); next, the child will overgeneralize inversion, essentially treating English as an INFL2 language, as dictated by (b.); and finally, the English setting (c.) will be attained.

5.12 Other Sources of Parametric Variation

The next parameter, the AGR-in-COMP parameter, is only of relevance to S-VP-INFL languages; assuming that the child at the relevant stage of acquisition has identified English as an S-INFL-VP language, we may conclude that AGR-in-COMP is irrelevant to the acquyisition of English. We therefore predict that children learning English will fail at any stage to produce structures associated with AGR-in-COMP, in particular, NCC structures.

Similarly, the embedded WH-adjunction parameter, dependent on the existence of marked complements of the type [-WH, +Q] - which must be learnt on the basis of positive evidence - will be inapplicable to English, which does not contain such complements.

This leaves two sources of further parametric variation. The first concerns the notion of "possible A'-type P-subject", as discussed in 4.54 above. Here we made no claim about

markedness, merely listing the relevant possible subclasses, with the possible exception of an implicational relationship between interrogative and exclamative P-subjects; the latter predicts that no child learning English will apply SAI incorrectly to structures containing exclamative WH-phrases unless s/he already applies SAI to at least some interrogative WH-phrases.

Finally, we turn to the dependent/independent clause parameter. "Absolute" predictions are hard to make here, since we are dealing with a continuous and not a discrete parameter. However, we have proposed a series of implicational statements, repeated below as (2), which allow us to make "relativized" predictions:

(2)a. If X/d, then X/i;
b. If X/i, then X/e,

where X is a variable over "non-canonical clausal type", d = declarative, i = interrogative, and e =exclamative.

On the basis of (2),we can predict that a child learning English will only produce dependent declaratives in matrix environments if s/he also produces dependent interrogatives in such environments, and dependent interrogatives in matrix environments just in case s/he also produces dependent exclamatives in such environments (presuming of course, that at the relevant stage of language acquisition, the child employs interrogative and exclamative clauses). Precisely the same reasoning applies to independent clausal types in subordinate

environments.

5.13 Summary

Summarizing, we make the following set of predictions for the acquisition of English, based on the parameters we have proposed, and a simple, linear-parametric model of language acquisition:

(3)a. Children learning English will pass through two stages in the acquisition of SAI, one involving no inversion, the other involving an "INFL2"-type stage.

b. Children learning English will never make NCC-type errors.

c. Children learning English will not apply SAI to exclamatives unless they do so with interrogatives.

d. Children learning English will only produce declarative non-canonical clausal types if they produce interrogative non-canonical clausal types, and interrogative non-canonical clausal types if they also produce exclamative non-canonical clausal types.

Let us then proceed to examine the data.

5.2 Overview of the Acquisition of WH-movement

5.20 Introduction

There are two preconditions to the acquisition of SAI: the acquisition of auxiliary verbs, which we have already investigated in Chapter 3, and the acquisition of SAI triggers, in particular WH-phrases, which we will now proceed to examine.

5.21 Early Acquisition of WH-Phrases

WH-phrases begin to appear in children's speech at the earliest stages of syntactic development. Indeed, in Wells' (1985) longtitudinal, multi-subject study, children as young as 15 months actually produce more WH-interrogative than declarative utterances; these consist of the single-word WH-questions what and where. At the two word stage, multi-word WH-interrogatives begin to appear; however, there is evidence that these first productions - consisting of ((WH-phrase + optional contracted copula) + NP) are at least partially rote-learned formulas, rather than cases of productive WH-movement. This evidence is provided by the following considerations:

(a) At the stage in question, children do not generally have productive control of the English auxiliary system (see Chapter 3 above). Often, the only instances of auxiliaries in the speech of children at this stage are those consisting of either

a WH-phrase or here/there followed by a contracted copula. In Wells' study, (where/what + copula) constructions begin to appear at eighteen months, six months before the first use of fronted auxiliaries in yes/no questions. Brown (1973) also reports that contracted forms of be are acquired after uncontracted forms (though see deVilliers and deVilliers 1973 for a different view, based on a cross-sectional study, and Kuczaj 1979 for yet a third opinion based on both longtitudinal and cross-sectional studies); if Brown's results are correct, then we have further grounds to view very early uses of WH + "contracted copula" as unanalyzed routines.

(b) There is a good deal of evidence that children at the stage in question are prone to segmentation errors involving contracted copulas. This evidence is of two main types. Firstly, children who in general have learnt the co-occurrence and ordering restrictions between copula be and other auxiliary elements frequently fail to observe them just in case a contracted copula is involved: Brown (1973), for example, reports that Adam (one of the three principal subjects of the Harvard longtitudinal study) failed to segment it's, on the basis of productions such as It's will go. See also Kuczaj (1976) and McWhinney (1981) for other examples of segmentation errors of this type. Second, WH-words with and without a contracted copula appear to be in free variation in the speech of many children at the relevant stage, indicating that the function of the auxiliary has not been grasped.

(c) As observed by Johnson (1981), (WH-phrase + contracted copula + XP) sequences are often used in a formulaic manner, sometimes in semi-ritualized naming games (e.g., "What's that ?... That's a...") which are predictable from context; see Ninio and Bruner (1978). Again, this is consonant with a rote-learned, routinized use of WH-phrases more comparable to idiomatic expressions such as "How're you doing ?" than to productive employment of WH-movement.

We therefore conclude that whereas WH-phrases are present from almost the earliest utterances of young children, the syntactic principles associated with their appearance are not present at these stages, at least as far as production data is concerned.

5.22 Acquisition of "Adult-Like" WH-Movement

The question then arises as to <u>when</u> the child begins to learn the adult system of WH-movement - or rather, when the investigator of language acquisition can <u>tell</u> when the child has learnt or is learning the adult system.

In order to answer this question, let us briefly review the evidence for a syntactic rule of WH-movement for adults. By far the most compelling evidence in monoclausal structures (which, after all, comprise the vast majority of children's utterances at the relevant stage of acquisition) comes from the existence of a filler-gap dependency. Note that we need say nothing about

transformations, movement rules, or any of the specific apparatus of contemporary generative theories in order to make this claim; any serious syntactic theory must relate the WH-word in sentence-initial position to a sentence-internal position in order to successfully interpret a WH-question. Now, it can be claimed that formulaic utterances such as what's that do not contain a filler-gap dependency: they can easily be construed as indexical expressions in the same way as There's a cat or That's mommy (for more discussion of this issue, see Peters 1983). Likewise, question forms such as the following, produced by Adam at MLU stage III, prior to productive use of WH-questions, are probably best treated as formulaic - a point made by Brown (1968):

(4)a. Why not he take bath ?
   b. Why not me careful ?
   c. Why not you looking right place ?

Here, there is no filler-gap relation between the sentence-initial WH-phrase and a sentence-internal position, and therefore no evidence for an adult-like rule of WH-movement. (Note also that there is in fact a quite separate model for the sentences in (4): the why not + imperative construction discussed in the last chapter, which shares none of the characteristic properties of WH-movement in English.)

Yet another example of a formulaic type of WH-question is presented by Labov and Labov (1978). Their daughter Jessie, the

subject of an intensive single-subject study, began to produce the following types of question at the age of twenty six months - again, before productive control of either the WH-system or the auxiliary system:

(5)a. How 'bout that, mama ?

b. How 'bout the wash ?

c. How 'bout a baby ?

Once again, the WH-expression How 'bout is clearly formulaic and gives no evidence of adult-like WH-structures. Interestingly, this type of How 'bout question persisted in Jessie's language at least until the age of 40 months, well after she had begin to acquire the adult structures. Thus we get the following examples, which show a considerable increase in overall linguistic maturity but exactly the same use of How 'bout:

(6)a. How 'bout you move so we both can have some space for lie down ?

b. How 'bout I move your hair out of the way: it's really hard for brush it.

What (6) shows is that early, routinized use of formulaic WH-expressions can persist well into advanced stages of acquisition; while perhaps not terribly surprising, this type of evidence does indicate the need for methodological vigilance not only in the early stages of acquisition but also at later

stages when - at least superficially - the child has grasped the essentials of the adult system.

It is also important to note that in English, neither subject questions nor adjunct questions provide evidence for filler-gap dependencies, either, since such structures contain no uniquely identifiable gap. Thus the folowing questions, produced by the Harvard children at MLU stage III, provide no evidence for an adult-like rule of WH-movement:

(7)a. What soldier marching ?

b. Where my mitten ?

c. Why you waking me up ?

None of these structures contain an overt filler-gap dependency; (7a.) and (b.) could be analyzed as (WH + NP) in the same way as an indexical expression such as What's that ?, and (6c.) might well be a why question of the type in (4) - i.e., (why(not) + S).

Having now examined three types of construction which provide no evidence for adult-like WH-movement, it should be obvious what kind of structure does provide the right kind of evidence. Recall that in the adult system, WH-questions involve a relationship between a sentence-internal argument or adjunct position and a sentence-external "operator" position. Thus as soon as the child begins to use (and successfully interpret) WH-questions containing overt filler-gap dependencies, we can conclude that the essence of the system

has been grasped. Only one type of WH-extraction contains unequivocal, overt gaps: that involving extraction of a subcategorized complement. We can therefore conclude that an adult-type WH-system has been acquired from examples such as the following, again taken from the Harvard corpus, this time at MLU stage IV:

(8)a. What I did yesterday ?

b. Where he can put them ?

As long as the child does not employ the relevant verbs only with S-initial WH-phrases and a gap, or only with a gap (neither of which is very plausible with such common verbs as do and put), we can safely conclude that the examples in (8) are prima facie evidence for an adult-like system of WH-"movement".

We thus conclude that the child has an "adult-like" rule of WH-"movement" as soon as complement WH-phrases appear with verbs which are already used (with the appropriate number of arguments) without WH-phrases. The age at which this stage is reached appears to vary widely between children, not a surprising result in view of the often observed discrepancy between chronological age and various measures of linguistic development (on this point, see in particular Cazden 1972, Brown 1973, and Wells 1986). In terms of developmental sequencing, however, there does appear to be an interesting correlation between acquisition of productive WH-movement and

acquisition of the auxiliary system: both appear to come in around MLU stage IV.

It is important to note that this evidence is entirely independent of the question of whether WH-words are "base-generated" or "moved" in the grammars of young children: indeed, that whole question might turn out to be vacuous, given the thrust of current generative theory towards conditions on representation rather than conditions on derivation. This renders irrelevant the frequently made claim that since young children never produce WH-phrases in situ, there is no evidence that WH-movement ever takes place (see for example Klein 1982, Pinker 1984, O'Grady 1987, to name a few advocates of this argument).

While it is difficult to establish this correlation more precisely - as indicated in Chapter 1, this is a perennial problem with evidence based on acquisition sequences - it is highly suggestive, in that it indicates that two of the most important clausal syntactic subsystems are acquired at approximately the same time. From either a "maturational" or a "constructivist" view of language acquisition this is interesting; from the former viewpoint, it could be taken to indicate that both the "WH-subsystem" and the "auxiliary subsystem" are determined by the same type of maturation; from the constructivist viewpoint, that both require the same linguistic and non-linguistic preconditions for their emergence.

5.23 Lexical Specificity in the Acquisition of WH-Movement

Before turning directly to the interaction of these two components - i.e., the acquisition of SAI - one other important aspect of the acquisition of WH-movement should be discussed. This is the "lexical specificity" of the acquisition of WH-words. As documented by numerous studies, including those of Wells (1985), WH-words are acquired one at a time, in a remarkably consistent sequence; the following table, taken from Davis (1983) combines the results of several published studies as well as an examination of unpublished corpora:

(9) Order of acquisition of WH-words:

(a) What/Where
(b) Why/how
(c) Who
(d) Which X/ How X

What is the reason for this developmental sequence ? It appears unlikely that any one factor can account for the observed pattern; rather, it seems probable that at least the following variables influence the sequence of acquisition:

(a) Frequency in the input. Though the relative importance of "token frequency" and "type frequency" is still quite unclear, it is reasonable to suppose that one, or both, have a

significant effect on the acquisition of lexical items. In the case of WH-words, it appears likely that what and where, which typically involve the location and identification of objects in the "here and now", are more frequent in discourse directed at young children than either why or how, which involve inferential reasoning, or when, which involves temporal sequencing often beyond the capabilities of three or four year olds. This leaves who, which is acquired surprisingly, but consistently, late; Wells (1985) speculates that this is due to frequency factors, but without a quantative study of the proportion of each WH-word in a representative (and large) sample of discourse to young children, this must remain a speculation.

(b) Cognitive difficulty. Children of two and three have no problems with identification or location of objects (in Piagetian terms, they have passed the stage of object permanence) but they do have difficulty with temporal sequences, for example, which typically involve "temporal decentration" and "reversability" (again, resorting to Piagetian terminology), both levels of conceptual development not attained until the ages of at least three and four, and possibly later (for a review of the relevant literature, see Weist 1986). This in part surely explains the late acquisition of the temporal WH-word when. Similar considerations apply to how and why, both of which involve notions of causality often not available to young children. (Note, however, that a full

understanding of the concepts underlying a given lexical item is not necessarily a prerequisite for the use of that item; young children often use why questions, for example, as conversational gambits rather than in a genuine information-seeking capacity). Of course, it is very difficult to disentangle the effects of cognitive difficulty from those of frequency in the input, especially since, as often noted, discourse to children is "fine-tuned" to the appropriate cognitive level of the child. It is no coincidence, then, that the "input frequency" and "cognitive difficulty" factors make virtually the same predictions concerning acquisition sequence (which, incidentally, means that who is anomalous on both accounts). Nevertheless, explanations for language acquisition sequences based on "cognitive complexity" do make certain predictions; for example, if the late acquisition of when is due to cognitive complexity, we should not expect to find children who fail to employ when successfully using other temporal conjunctions. It is beyond the scope of this thesis to investigate such predictions thoroughly; there is a vast literature on the relation between linguistic and cognitive development, which I will not attempt to review here.

(d) Properties of the learning mechanism. We have advocated throughout this thesis a distinction between "open class" and "closed class" learning. We have proposed that the latter is mediated by a Parallel Distributed Processsing mechanism of the type described in McLelland and Rumelhart (1986), which

predicts a particular set of properties in a developmental sequence: sensitivity to input frequency, type- and token-variability, and "probabilistic" acquisition. Since WH-phrases constitute a closed class, the fact that they are acquired in an item-by-item manner falls out straightforwardly from this type of learning.

We conclude that extra-grammatical factors are mainly or entirely responsible for the pattern in (9). In fact, it should be noted that linguistic difficulty is probably the least likely contributing factor. The only syntactic distinction between WH-words which might be relevant is that between arguments and adjuncts. However, this distinction conspicuously fails to predict the order of acquisition of WH-words: while what, which is acquired early, is invariably an argument, where, which is also acquired early, is just as likely to be an adjunct; and whereas when, an adjunct, is acquired late, who, an argument, is acquired even later. We conclude that syntactic theory has virtually nothing to say about the order of acquisition of WH-words.

This negative finding is not without interest for the linguist attempting to come to grips with data from language acquisition, however, for two reasons. First of all, the item-by-item pattern of acquisition discussed above is exactly what we should expect for the acquisition of a closed-class grammatical category, regardless of the specific order in which the items are learnt. Thus the acquisition of WH-words supports

our contention - first made in Chapter 2 - that WH-phrases are G-heads, and at the same time appears to confirm once again our prediction, based on the Designation Convention, that closed-class items should be acquired "lexically" rather than "categorically". Second, the fact that syntactic factors have no influence on the actual order of acquisition of WH-words is actually a positive finding for the developmental psycholinguist. This is because one of the most urgent problems in the theoretical approach to language acquisition data is in knowing what not to explain; since inevitably, the data itself is a complex tangle of contributing factors, this problem can only be really resolved by a concerted effort on all fronts to clarify the respective domains of the many - presently competing - approaches to language acquisition, whether they be linguistic, cognitive, or psycho-social. By relinquishing the need to explain certain aspects of developmental data, the linguist is indirectly contributing to such a clarification.

Let us then summarize this section. We have seen that, after a period of "routinized" use of formulaic expressions involving WH-phrases, the child typically begins to acquire an adult pattern of WH-movement (though obviously, one restricted to single clauses) around MLU stage IV, at approximately the same time as the acquisition of an adult-like auxiliary system. We have also seen that the development of the WH-system, just like the development of the auxiliary system, proceeds lexical item by lexical item, giving further support to the contention that both constitute "grammatical" (G-head) classes rather than

"lexical" (θ-head) classes. We have now set the stage for the interaction of these two classes - in other words, the acquisition of SAI itself.

5.3 The Acquisition of SAI: Longtitudinal Studies

5.30 Introduction

The acquisition of SAI in English is dominated by a very famous set of "undergeneralization" errors involving the failure of inversion in WH-questions. Though intensively studied ever since Bellugi's pioneering work on the acquisition of interrogative structures (see Klima and Bellugi 1966, Bellugi 1971, Cazden 1972) both the extent and the nature of these errors is still under dispute. We will begin our discussion, then, with a presentation of Bellugi's work, and go on to examine the grammars of three other children whose acquisition of SAI, as we shall, differed rather strikingly from that reported by Bellugi.

5.31 The Harvard children

Employing longtitudinal data from the three children in the Harvard study, Bellugi distinguished six "periods" in the acquisition of questions. These are reproduced below in (10) (modified from the tables in Cazden 1972, p.54, with supplemental data from Bellugi 1971):

(10)

PERIOD A

AGE : 27-35 months (Adam); 27-36 months (Sarah); 19-22 months (Eve).
MLU : Stage I - II

Auxiliaries: none (excluding unanalyzed contracted copula). Yes/no questions signalled exclusively by intonation: e.g. Sit chair?, Ball go?
WH-phrases : restricted to formulaic routines: e.g What('s) that?, Where NP go?
SAI : none.

PERIOD B

Age : 35-36 months (Adam); 36-41 months (Sarah); 22-25 months (Eve).
MLU : Stage III

Auxiliaries: appearance sporadically of first negative auxiliaries can't and don't. Yes/no questions still signalled exclusively by intonation: e.g. Mom pinch finger?, You can't fix it?.
WH-phrases : More sophisticated WH-questions, but still no evidence of adult WH-system. Subject WH-questions appear: e.g, Why you waking me up?, What soldier marching?
SAI : none.

PERIOD C

Age : 36-42 months (Adam); 41-43 months (Sarah); 25-27 months (Eve).
MLU : Stage IV

Auxiliaries: rapid growth in use of auxiliaries. Inverted auxiliaries appear in yes/no questions, first in positive, then in negative questions: e.g., Are you going to make it with me?
WH-phrases : "true" adult-like WH-questions appear: e.g., What I did yesterday ?, Will you help me? Appearance of first embedded WH-constructions: I used to wear that when I little baby, That what I do.
SAI : SAI in yes/no but not WH-questions.

PERIODS D-F

Age : 42 months plus (Adam); 43 months plus (Sarah); 27 months plus (Eve).

MLU : Stage V+

Auxiliaries: Development of tag questions, beginning with huh in periods D and E: e.g., I have two turn, huh ?, We're playing, huh? followed by tags with inversion in periods E and F: e.g., That's funny, isn't it?, He was scared, wasn't he?

WH-phrases : Developmemt of complement WH-phrases (in period F): e.g., You don't know where you're going., He doesn't know what to do.

SAI : in periods D and E, inversion in yes/no questions and in positive WH-questions, but not in negative WH-questions, e.g., Why are you thirsty ?, Why we can't find the right one ?; in period F, inversion in negative WH-questions as well: e.g., Why can't they put on their diving suits and swim ? Occasional (one out of five in Adam's sample) inversion in embedded WH-complements: e.g, I don't know what are dey?

The following generalizations emerge from the data presented in (10):

(a) Bellugi's "periods" seem to correlate quite closely with MLU, but not very closely with age.

(b) The emergence of SAI coincides with the emergence of both adult-like auxiliary- and WH-systems.

(c) Children learn SAI in an invariant sequence.

(d) The developmental sequence of SAI is affected by two parameters: WH-question versus yes/no question, and negative question versus affirmative question.

(e) Tags are learnt later than matrix questions.

(f) Embedded questions are learnt later than matrix questions: there is a slight tendency to invert in indirect questions.

It should be noted that some parts of the empirical foundation on which these generalizations rest has been questioned by many subsequent researchers. Thus before turning to a discussion of theoretical implications, we will give an overview of the debate that has ensued over the data itself, dealing with the empirical generalizations in (a.-f.) one by one.

Few would question generalization (a.). It is well-known that age is a poor index of syntactic development (see our comments above), and as for MLU, it is unsurprising that a correlation should exist between certain types of syntactic development (for example, the acquisition of auxiliary verbs) and length of utterance, given that sentences with auxiliaries (which are obligatory in interrogative environments) are longer than those without. However, even MLU is at best a weak indicator of syntactic development, as we shall see in our investigation of the SAI acquisition of K., whose grammar changed while his MLU remained fairly constant.

Generalization (b.), likewise, is uncontroversial. Given that SAI is dependent on (a) auxiliaries and (b) SAI-triggers, the acquisition of SAI must await the appearance of both.

Generalization (c.), however, is far from uncontroversial, and has been the focus of considerable attention in the literature. What is at issue is the status of period C, the

extent of variation between children, and the reasons for such variation. Bellugi's claim, made on the basis of the three Harvard children, is that period C, though obviously variable in length, characterizes the "intermediate" grammars of all children learning English as a first language; while she admits that "Adam gave us the clearest picture of consistency and change" (Bellugi 1971, p.101), she attributes the sequence in (10) to all the children she investigated. It should be noted, however, that even Adam in period C did produce some WH-questions where SAI applied. The proportion of non-inverted to inverted WH-questions with modal auxiliaries in Adam's corpus are reproduced here from Bellugi (1971):

(11)

	Affirmative		Negative	
Samples	Non-inverted	Inverted	Non-inverted	Inverted
1-16	-	-	-	-
17-24	3	-	2	-
25-34	22	8	9	-
35-42	5	33	5	-
43-51	4	27	-	5

(Sample nos. 1-5 correspond to period A; 17-19 to period B; 24-26 to period C; 32-34 period D; 40-42 period E; and 49-51 period F.)

Though the pattern in (11) by and large bears out Bellugi's claim, note that it was only in period B, when neither the auxiliary system nor the WH-system had been acquired, that Adam produced only uninverted WH-questions. On the other hand, during period C, 20.5% of his WH-questions were inverted; though that proportion dramatically increased to 76.5% in

period D, and finally to 89% in period F, it appears that even the "clear and consistent" changes in Adam's grammars are not completely categorical.

In the years since Bellugi's work, it has become clear that in fact there is a rather wide range of variation in the proportion of inversion in children's questions (both WH-questions and yes/no questions). In the next two sections, we will be examining data from two children who show rather different developmental patterns from that described by Bellugi.

5.32 The child S.

There are some children who appear to go straight from period B to period F. That is, such children fail to go through any intermediate "non-inverting" stage; instead, they go from a "pre-auxiliary stage" to an adult-like rule of SAI. In this section, we will examine a child, S., whose spontaneous production was recorded at approximately bi-monthly intervals from the age of 18 to 47 months (thanks to Ken Reeder for supplying the data). S. provides an excellent example of a child who fails to pass through the non-inverting stage. The following table is an attempt to provide "periods" for S that can be compared with those postulated by Bellugi:

(12)

PERIOD A:

AGE : 20-24 months.
MLU : Stages I-II

Auxiliaries : Sporadic use of can't and don't: e.g., I can't find it, Don't wanna, daddy Yes/no questions marked by intonation only, e.g., Do S's hair ?, Wanna chocolate ?
WH-phrases : Confined to formulaic what and where questions, e.g. Where my doggy ?, What is it ?
SAI : Occasional, probably formulaic: Would you let me out ?

PERIOD B:

AGE : 24-30 months.
MLU : Stage II

Auxiliaries : appear in both declaratives and yes/no questions: e.g., You must have...find my bowl, Oh, he might go and sing, Can you go up there ?, Will you do dat one well ?
WH-phrases : adult-like WH-movement, but no auxiliaries except for unanalyzed contracted copula: What you gonna wear ?, What's he's got ?, Where's dad go ? First embedded Wh-phrases, e.g., Dis is where green parrot flying , Look what I got, all blood
SAI : in most yes/no questions, with a couple of exceptions, e.g., Dad did come outside ?; never in WH-questions.

PERIOD C:

AGE : 30-36 months.
MLU : Stages II-V

Auxiliaries : used correctly in all environments, including declaratives, yes/no questions, and WH-questions. However, still a tendency to omit in WH-questions: Where the strawberries gone away ?, What dis one called ? First tags appear, without auxiliaries: Dad, you get a very better big toe, yes ?,
WH-phrases : adult-like WH-movement, but still a lot of WH-questions withoout auxiliaries: What you say ?, Where daddy get cough cough ? Increase in embedded WH-constructions, including adjunct phrases: You mean, when I finish, I get a big bit, yes ?, Dad, I'm putting it back where I'm storing it
SAI : in all yes/no questions, most WH-questions with auxiliaries, e.g., Where are you, raggedy ?, Where did you go ?; however, fails to apply in

some WH-questions, e.g., What you said ?, What he boughted again ?

PERIOD D:

AGE: 36 months plus.
MLU: Stage V plus.

Auxiliaries: appearance of polar tags with auxiliaries: I live on through dat, isn't that right ?, Haven't got enough, have ya ? WH-questions without auxiliaries disappear.
WH-phrases : appearance of relative clauses, and other complex WH-constructions: And that's the way the snake's sleeping, Whenever we go to parks and things.... First extraction from embedded clause: What did he want to know ?. First WH-exclamative: Hey, how high it is, look at that !
SAI : applies in all adult environments, except for negative WH-questions, e.g., What you aren't writing it ?

The data summarized in (12) contrasts with that of Bellugi's study in a number of ways, but before we go on to discuss these contrasts, let us first take note of some similarities. The acquisition of the auxiliary system (with can't and don't preceding the other auxiliaries) mirrors the sequence reported by Bellugi. In addition, the acquisition of tag-questions follows the pattern described by Bellugi, with a non-polar tag marked by rising intonation preceding a full polar tag containing an auxiliary element. Finally, embedded WH-clauses appear at the same stage of S.'s development (period B) as in Bellugi's study (period C); while this might not appear to be the case on the basis of MLU, these two periods share a set of closely correlated syntactic milestones: acquisition of the auxiliary system, acquisition of WH-movement, and acquisition of embedded WH-phrases.

The first important difference between the two sets of data is in the correlation of syntactic development with MLU. Whereas in all of Bellugi's subjects, MLU stage IV marked the acquisition of the auxiliary and WH-systems, S. appears to pass through the same stage of syntactic acquisition at a time when her MLU was only at stage II. Subsequently (in period C) her MLU rose very rapidly, but it is clear that MLU and syntactic acquisition are not as closely linked as generalization (a), made on the basis of the Harvard data, would indicate.

Let us now turn to generalizations (c) and (d). Recall that Bellugi's data indicated a stage (her period C) when inversion took place in yes/no questions but not in WH-questions. The corresponding period (B) in S.'s development, however, yields a rather different picture. While there <u>is</u> a clear discrepancy between the acquisition of yes/no questions (which follow the pattern described by Bellugi) and WH-questions, this discrepancy takes a rather different form: WH-questions are consistently produced <u>without auxiliaries</u>, while yes/no questions are produced with auxiliaries and inversion. In (13) below, S.'s WH-questions without auxiliaries, with contracted auxiliaries, and with uncontracted auxiliaries are detailed for periods A, B, and C:

(13)

	WH -AUX	WH +CON. AUX	WH +UNCON. AUX
Period A:	2	-	3
Period B:	23	19	1
Period C:	18	41	38

Note, however, that S. did produce some WH-questions with auxiliaries and without inversion - not in period B (which corresponds to period C in Bellugi's account), but rather in period C (which corresponds to period D in Bellugi's account). The proportion of inverted and non-inverted WH-questions with uncontracted auxiliaries during S.'s periods C and D are detailed below in (14):

(14)

	Affirmative		Negative	
	Non-inverted	Inverted	Non-inverted	Inverted
Period C	3	35	0	0
Period D	0	28	1	1

Two aspects of (14) are notable. First of all, the proportion of non-inverted to inverted WH-questions is very low even in period D; this seems to show that S. never really went through a stage comparable to Bellugi's period C. Second, S., unlike Adam - see (11) above - produced a vanishingly small number of negative WH-questions; this prevents us from asssessing Bellugi's claim concerning the diferential acquisition of SAI

in negative and affirmative WH-questions.

Thus our comparison between the Harvard children and S. shows that at least some of the generalizations made on the basis of Bellugi's work cannot be maintained. In the next section, we will review another developmental pattern, that of a child who did pass through a "non-inversion" stage, but in a rather different manner than the Harvard children.

5.33 The Child K.

In this section, we will be examining the SAI development of a child, K., whose grammar I studied intensively for a period of about six months, when he was from 33 to 40 months old. While this comparatively short time span prevents us from comparing K.'s long-term development to that of S. or the Harvard children, my investigation of K.'s grammar does afford a rather more intensive look at a particular developmental stage (the "non-inverting stage") which I believe can reveal interesting sub-regularities that are often lost in the broad sweep of a more general longtitudinal study.

During the course of the investigation, I employed a number of experimental techniques designed to probe K.'s grammatical competence more deeply than would have been possible on the basis of data from spontaneous production alone. These included act-out tasks, elicited production techniques, and imitation techniques. For the purposes of the investigation of SAI, elicited imitation proved particularly valuable, and I will

present imitation data along with data from spontaneous production in my examination of K.'s SAI development. Unfortunately, this supplemental experimental data is not as methodical as I would have liked, since my investigation of K.'s grammar was exploratory rather than systematic; nevertheless, it seems to me that this pilot study represents a promising approach to the investigation of syntactic acquisition, which has the potential to reveal a much more fine-grained picture of developmental sequences than has previously been possible.

I have divided the six month time-span of the investigation into five equal "periods" of approximately a month each; this is purely in the interests of data presentation, and embodies no claims about developmental "stages". Imitation data will be presented for periods II - IV alone, since for various reasons not enough relevant imitations were produced in the other two periods. It should be noted that K.'s MLU during the entire six-month span of the investigation was between 3.3 and 3.8; this places him in MLU stage IV (equivalent to Bellugi's period C) throughout the developmental sequence which we are about to examine.

Let us accordingly turn to Period I, when K. was 33-34 months old and his MLU averaged 3.56. The table below summarizes his spontaneous production of questions during this period:

(15)

PERIOD I

Matrix questions

	O	I	U	U/C	C	S	TOTAL
What	20	0	6	3	21	0	50
What X	0	0	1	0	0	0	1
Where	1	0	0	0	5	-	6
Why	0	0	2	0	0	-	2
Who	0	0	0	0	0	1	1
TOTAL WH	21	0	9	3	26	1	60
YES-NO	19	0	1	-	-	-	20
TOTAL Q	40	0	10	3	26	1	80

Embedded questions

	O	I	U	U/C	C	S	TOTAL
What	4	0	4	0	0	0	8
Why	1	0	2	0	0	-	3
When	0	0	2	0	0	-	2
TOTAL WH	5	0	8	0	0	0	13
YES-NO	0	0	0	-	-	-	0
TOTAL Q	5	0	8	0	0	0	13

KEY: O = omitted I = inverted U = uninverted
C = contracted U/C = uninverted/contracted
S = subject

The most prominent characteristic of Period I in K.'s development is the number of omitted auxiliaries in his questions. Auxiliaries were omitted in fully half of all matrix

questions, and in 19 out of 20 yes-no questions. In fact, K. characteristically produced yes/no questions by dropping both subject and auxiliary at this period, as illustrated in (16) below:

(16)a. Want to hear that, you two guys ?
b. Want come to the hospital ?
c. See that funny couch ?

Such questions typically involved a small set of verbs, including want, remember, see, and know, and persisted throughout the study period. In a sense, they could be seen as an "avoidance strategy", providing a small set of sentence frames which could be employed in order to circumvent the difficulties of attempting to apply SAI in yes-no questions. However, in the same period K. also produced yes-no questions with normal declarative order and rising intonation, as in (17):

(17)a. You got any scramble eggs here ?
b. You want any more apples ?
c. You make a wowo bug ?

It should be noted that the high proportion of omitted auxiliaries in K.'s questions was not due to a more general tendency to omit auxiliaries; at this same period, K. usually employed auxiliaries correctly in declarative structures, as in

(18) below:

(18)a. We can do that one more time.

b. I can't do it.

c. I don't know.

In fact, in declarative sentences, K. employed auxiliaries in 75% of obligatory contexts; while this does not reach Brown's (1973) criterion of 90%, it is still considerably higher than the proportion employed in questions (50%, including auxiliaries contracted to WH-phrases, which, as we shall argue below, were not analyzed correctly by K. at this point; if we exclude the contracted cases, the percentage drops to 22.5%).

This finding conflicts with Bellugi's claim that auxiliaries emerge at the same time in declarative and yes/no questions (and also, incidentally, casts further doubt on Hyams' (1983) account of the "missing subject stage" in the acquisition of English, which relies heavily on Bellugi's account of auxiliary development: see Chapter 2 for details).

Turning to WH-questions, note that while K. produced no inverted auxiliaries at this period (excluding contractions), he did produce a small but significant percentage of WH-questions containing uninverted auxiliaries (21.5% of matrix WH-questions). Moreover, K. "de-inverted" WH-questions when asked to imitate them (numbers are too small to be significant, but the data is nonetheless suggestive):

(19) Target: What am I gonna do ?
Imitation: What I'm gonna do ?

Finally, let us turn to the question of contraction. There are several reasons to believe that in Period I K. failed to correctly analyze (WH + contracted auxiliary) sequences. The first is the complete absence of inverted auxiliaries in uncontracted environments during this period. The second is provided by sentences in which K. produced both a (WH + contracted auxiliary) sequence sentence-initially and an auxiliary in medial position; examples are given below:

(20)a. And what's the mouse is doing ?
b. What's he's doing ?

Moreover, such constructions appear to be in free variation with similar sentences in which the auxiliary is omitted altogether, or produced in medial position; the following examples were produced in the same session as those in (20):

(21)a. What the mouse is doing ?
b. What they're doing right now ?
c. What he saying ?
d. What he's saying ?

Third, K. produced contracted WH-questions in which the contracted auxiliary failed to agree with its subject, whereas in declarative environments subject-verb agreement was

generally employed correctly:

(22)a. What's these flat things ?

b. Where's the eyes ?

c. What's these for ?

Finally, there is evidence that K. misanalyzed other (NP + contracted auxiliary) sequences, notably they're, as shown by the examples in (23):

(23)a. They're say yeah.

b. They're got dinosaurs.

c. No they're don't.

We conclude that in Period I K. has no rule of SAI at all, even though there is evidence that he controls auxiliaries in declarative environments; auxiliaries contracted to WH-phrases are not correctly analyzed at this point.

Briefly turning to K.'s embedded questions, we see that more or less the same pattern holds as in matrix questions, with one exception: there are no (WH + contracted auxiliary) sequences in embedded clauses. This is interesting, in that it shows that while K. has not correctly analyzed contracted auxiliaries, he does not simply treat them as phonological variants of WH-phrases; if he did, we would expect to find them equally frequently in embedded clauses. It appears that K. has misanalyzed contracted auxiliaries, rather than failed to

analyze them altogether.

Let us then turn to Period II, when K. was 34-35 months old, and his MLU average was 3.54. Once more, we begin by presenting his spontaneous question-production:

(24)

PERIOD II

Matrix questions

	O	I	U	U/C	C	S	TOTAL
What	5	3	3	9	49	3	72
What X	1	0	0	0	0	0	1
Where	7	2	0	0	6	-	15
Why	2	1	5	0	0	-	8
How	1	0	0	0	0	-	1
How X	0	1	0	0	0	-	1
Who	0	0	0	0	1	0	1
TOTAL WH	16	7	8	9	56	3	99
YES-NO	18	0	2	-	-	-	20
TOTAL Q	34	7	10	9	56	3	119

Embedded questions

	O	I	U	U/C	C	S	TOTAL
What	0	0	1	0	4	0	5
Why	1	0	1	0	0	-	2
TOTAL WH	1	0	2	0	4	0	7
YES-NO	0	0	0	-	-	-	0
TOTAL Q	1	0	2	0	4	0	7

KEY: O = omitted I = inverted U = uninverted
C = contracted U/C = uninverted/contracted
S = subject

There is one striking and obvious difference between Periods I and II: while the number of WH-questions with omitted auxiliaries has dropped significantly, the proportion with contracted auxiliaries has risen dramatically, and has even been overgeneralized to embedded clauses. In contrast, yes-no questions in Period II show an almost identical pattern to that of Period I, being chiefly characterized by omission of the auxiliary. The significance of this pattern will be discussed in 5.7.

Question-imitation data for K. during period II is presented below:

(25) QUESTION IMITATIONS: PERIOD II

Matrix questions

		YES-NO		WH		TOTAL
		I	U	I	U	
Target		3	0	3	1	7
Imitations	I	0	-	0	0	0
	U	1	-	0	0	1
	O	2	-	3	1	6

Embedded questions

		YES-NO		WH		TOTAL
		I	U	I	U	
Target		0	10	2	4	16
Imitations	I	-	0	0	0	0
	U	-	9	0	0	9
	O	-	1	2	4	7

KEY: I = inverted, U = uninverted, O = omitted.

While imitation data at this stage is rather sparse, the results generally seem to follow the pattern of K.'s spontaneous production of questions; inverted questions are imitated either without an auxiliary (usually) or "de-inverted" (occasionally). One surprising aspect of the imitation data was K.'s treatment of the WH-complementizer if/whether. When presented with embedded yes/no questions, K. consistently substituted a WH-phrase for the complementizer, as in the following examples:

(26)a. Target: Do you know if the mouse fell off the ladder ?
Imitation: Know how mouse fell off the ladder ?

b. Target: I don't know if the paintbrush fell off.
Imitation: Don't know what the paintbrush fell off.

c. Target: The water overflows if you leave the tap on.
Imitation: The water overflows when you leave the tap on.

d. Target: I don't know whether it's Scuffy.
Imitation: I don't where is Scuffy.

It appears that K. used an "avoidance strategy" in embedded yes/no questions, just as he did in matrix yes/no questions, though it is difficult to say exactly why; it may have been that he simply had not yet acquired <u>if</u>, in which case his only way of formulating embedded yes-no questions would be by supplying a WH-phrase instead. Note that at this same period - and indeed, throughout the time-span during which I investigated his grammatical development - K. consistently deleted the complementizer <u>that</u> in imitations of all types, including relative clauses, declarative complements, and clefts. Examples are given in (30) below:

(30)a. Target: I saw the cow that knocked over the milk.
Imitation: I saw the cow knock over his milk.

b. Target: I think that she'll come in 15 minutes.
Imitation: I think she gonna come in 15 minutes.

c. Target: It's funny that a goat is lying down.
Imitation: That's funny the goat's lying down.

Deletion of <u>that</u> was not triggered by a general failure to imitate these constructional types; as shown in (31) below, K.

more or less successfully imitated WH-relatives, declarative complements without a complementizer, and <u>what</u>-clefts:

(31)a. Target: This is a girl who has a neck.
Imitation: Now this is a girl what has a neck.

b. Target: I think she'll come soon.
Imitation: I think she will come now.

c. Target: What the old lady needs is some food.
Imitation: That what the lady needs some food.

Neither was deletion of the <u>that</u> complementizer tied to lexical properties of <u>that</u>; K. used <u>that</u> as a determiner and deictic pronoun quite freely during the period when he consistently refused to employ it as a complementizer.

Let us then turn to Period III, when K. was 35-36 months old and his MLU averaged 3.32. Once again, we begin with spontaneous question production:

(32)

PERIOD III

<u>Matrix questions</u>

	O	I	U	U/C	C	S	TOTAL
What	5	0	12	0	3	3	23
Where	0	0	2	0	1	-	3
Why	4	0	3	0	0	-	7
TOTAL WH	9	0	17	0	4	3	33
YES-NO	15	0	5	-	-	-	20
TOTAL Q	24	0	22	0	4	3	53

Embedded questions

	O	I	U	U/C	C	S	TOTAL
What	0	0	1	0	1	0	2
When	0	0	3	0	0	-	3
How	0	0	2	0	0	-	2
TOTAL WH	0	0	6	0	1	0	7
YES-NO	0	0	1	-	-	-	0
TOTAL Q	0	0	7	0	1	0	8

KEY: O = omitted I = inverted U = uninverted
C = contracted U/C = uninverted/contracted
S = subject

As dramatically as it rose in Period II, the proportion of contracted WH-questions has fallen in Period III. The dominant type of WH-question now contains an auxiliary in the uninverted position: i.e, K. is now in the "non-inverting" stage. Note, however, that K. still never inverts in yes-no questions; just as in periods I and II, he prefers to omit the auxiliary altogether, though the proportion of uninverted yes-no questions has also risen.

Let us then turn to imitation data:

(33) QUESTION IMITATIONS: PERIOD III

Matrix questions

		YES-NO		WH		TOTAL
		I	U	I	U	
Target		19	1	23	12	55
Imitations	I	3	0	11	3	17
	U	12	1	8	5	23
	O	4	0	4	4	12

Embedded questions

		YES-NO		WH		TOTAL
		I	U	I	U	
Target		0	2	4	6	12
Imitations	I	-	0	1	1	2
	U	-	2	2	4	8
	O	-	0	1	1	2

KEY: I = inverted, U = uninverted, O = omitted.

This data contrasts in an interesting manner with that in (32). Whereas the spontaneous production data showed a pattern of auxiliary omission in yes-no questions and uninverted auxiliaries in WH-questions, the imitation data show that while yes-no questions were mostly of the uninverted type, WH-questions were as likely to contain inverted auxiliaries. Indeed, K. even inverted three uninverted questions.

Note once again that, just as in periods I and II, K.'s yes/no questions appear to be <u>less</u> likely than his WH-questions to

contain inverted auxiliaries; of course, this is precisely the opposite finding from that reported by Bellugi.

During Period III, I also gave K. several "echo" questions to imitate. By and large, he imitated them correctly, though in one case he fronted the WH-phrase. Examples are given in (34):

(34) Target: You saw what yesterday ?
Imitation: What I saw yesterday ?

b. Target: Your mother knows who ?
Imitation: My mommy knows who ?

c. Target: You went to the store when ?
Imitation: I went to the store when ?

These examples show that K. at this period has no problem relating the "launching sites" to the "landing sites" of WH-movement; (34a.), in particular, indicates that he is fully aware of a syntactic chain linking S-external and S-internal WH-positions. This in turn indicates that there is no "deep" syntactic explanation for the absence of sentences containing in-situ WH-phrases in data from language acquisition; the most likely explanation is that they are rare in the input and only used in a very restricted set of real-world contexts.

I also gave K. some exclamatives to imitate during period III; interestingly enough, he did invert one of them, though unfortunately there are too few examples to tell whether this represented a real pattern. Examples are given below:

(35) Target: How tall the plants have grown !
Imitation: How tall the plants is growing !

b. Target: What a tall bridge this is !
Imitation: What a tall bridge is this !

c. Target: What a silly horse this is !
Imitation: This is... this a funny little silly horse !

Again, (35b.) might be a sign of "pre-productive competence" (see Kuczaj and Maratsos 1975), involving the overgeneneralization of SAI to an exclamative WH-phrase (as we shall see, children do overgeneralize SAI to how come questions); however, there is too little data to draw any other than speculative conclusions.

Let us then turn to Period IV. At this point, K. was 37-38 months old, and his MLU average was 3.38. His spontaneous question-production during this period is detailed below:

(36)

PERIOD IV

Matrix questions

	O	I	U	U/C	C	S	TOTAL
What	6	1	23	2	5	12	49
Where	0	0	2	0	2	-	4
Why	4	0	4	0	0	-	8
Who	0	0	0	0	0	2	2
How	4	0	0	0	0	-	4
TOTAL WH	14	1	29	2	7	14	67
YES-NO	15	1	2	-	-	-	18
TOTAL Q	29	2	31	2	7	14	85

Embedded questions

	O	I	U	U/C	C	S	TOTAL
What	0	0	2	0	0	1	3
When	0	0	2	0	0	-	2
How	1	0	1	0	0	-	2
TOTAL WH	1	0	5	0	0	1	7
YES-NO	0	0	4	-	-	-	4
TOTAL Q	1	0	9	0	0	1	11

KEY: O = omitted I = inverted U = uninverted
C = contracted U/C = uninverted/contracted
S = subject

The pattern exhibited here is very similar to that of Period III; WH-questions are mostly uninverted, yes-no questions omitted. Note the rise in subject WH-questions, and the appearance of embedded yes-no questions, which mark the acquisition of the WH-complementizer if.

Imitations for Period IV are provided below:

(37) QUESTION IMITATIONS: PERIOD IV

Matrix questions

		YES-NO		WH		TOTAL
		I	U	I	U	
Target		3	0	13	0	16
Imitations	I	0	-	0	-	0
	U	1	-	4	-	5
	O	2	-	9	-	11

Embedded questions

		YES-NO		WH		TOTAL
		I	U	I	U	
Target		2	3	1	8	14
Imitations	I	0	0	0	0	0
	U	2	3	1	8	14
	O	0	0	0	0	0

KEY: I = inverted, U = uninverted, O = omitted.

Again, the pattern in Period IV seems to be a consolidation of that established in Period III; note that this time, inversion takes place neither in WH-questions nor in yes-no questions.

Finally, we come to period V, when K. was 38-40 months old, and his MLU had risen slightly to 3.72. Spontaneous question production is detailed below:

(38)

PERIOD V

Matrix questions

	O	I	U	U/C	C	S	TOTAL
What	3	2	3	0	7	4	19
What X	0	0	1	0	0	0	1
Where	1	0	1	0	1	-	3
Why	1	1	2	0	0	-	4
Who	1	0	0	0	3	1	5
How	2	0	1	0	0	-	3
How much/ many	0	2	0	0	0	0	2
TOTAL WH	8	5	8	0	11	5	30
YES-NO	20	4	1	-	-	-	25
TOTAL Q	28	9	9	0	11	5	55

Embedded questions

	O	I	U	U/C	C	S	TOTAL
What	0	0	6	0	0	0	6
What X	0	0	1	0	0	0	1
Where	0	0	4	0	0	-	4
Who	0	1	0	0	0	0	1
How	0	0	6	0	0	-	6
When	0	0	4	0	0	-	4
TOTAL WH	0	1	21	0	0	0	22
YES-NO	0	0	4	-	-	-	4
TOTAL Q	0	1	25	0	0	0	26

KEY: O = omitted I = inverted U = uninverted
C = contracted U/C = uninverted/contracted
S = subject

Period V represents the beginning of a transitional period from the "non-inverting stage" towards the adult grammar. Note that uninverted WH-questions are no longer dominant in matrix environments, and K. has begun to invert in yes-no questions. There is also one overgeneralization of inversion to an embedded environment.

Unfortunately, not enough imitation data of the relevant type is available for this period, so we are unable to examine K.'s SAI development in more detail. I did however, employ imitation to investigate two other areas of his grammar: topicalization and multiple WH-questions.

Topicalization generally presented no problems for K., as

can be seen from the examples in (39):

(39)a. Target: This bicycle, a policeman rides.
Imitation: This...this bicycle, a policeman rides

b. Target: This car, a thief drives in.
Imitation: This car, a fief... a thief drives in.

c. Target: This bridge, the pig threw the cat off.
Imitation: This bridge, the pig threw the cat off.

d. Target: On top of the snake's hat, a little tiny ant dances.
Imitation: On the snake's hat a little ant dances.

Notice that K. shows no propensity to invert in topicalized structures. Of course, this could simply be a reflex of the non-inversion stage; however, since in Period V K. had begun to invert in both yes-no and WH-questions, in that case we might expect a sporadic pattern of inversion in topicalizations. In fact, K. inverted an auxiliary in none of the 10 topicalizations that he imitated.

Multiple WH-structures were also imitated successfully:

(39)a. Target: Who's going where ?
Imitation: Who's going <u>where</u> ?

b. Target: What is who doing ?
Imitation: What who's doing ?

c. Target: Where is who sitting ?
Imitation: Where who is sitting ?

Note, however, that the "superiority" violations in (39b.) and (c.) (see 2.8 for details) do not seem to disturb K, although in both cases he de-inverted an inverted auxiliary, showing that such examples are indeed being processed through his

grammatical system. This is intriguing, in view of the very remote possibility that superiority could ever be learned on the basis of input evidence; clearly, the topic merits future research.

This concludes our investigation of K.'s syntactic development. In spite of the fragmentary and incomplete nature of much of the data presented here, I hope to have shown that the methodological approach I have taken shows some potential for future research, as well as providing a more detailed picture of the "non-inversion" stage in the auxiliary development of one child.

Before leaving K.'s grammar, however, let us briefly detail the discrepancies between K.'s SAI-development and that of the Harvard children, as reported by Bellugi.

The most striking difference is in the relationship between WH- and yes-no questions. Whereas Bellugi's Period C is characterized by inversion in yes-no questions but not in WH-questions, K.'s development if anything shows the opposite pattern: during the non-inversion stage, yes-no questions were <u>least</u> likely to be inverted. As we shall see, it is the Bellugi pattern rather than K.'s which appears to be exceptional.

Our more detailed examination of the non-inversion stage in K.'s development also reveals sub-regularities not mentioned by Bellugi; in particular, we noted a strong upsurge followed by a sharp drop in the number of WH-questions produced with contracted auxiliaries; we will explore the implications of this developmental pattern in 5.7.

Before turning to cross-sectional studies, we will examine one more longtitudinal study of question acquisition, that of Labov and Labov (1978).

5.34 Jessie

The Labov and Labov study is based on a corpus of 3,368 WH-questions collected from their daughter Jessie from "the very first occurrence to the acquisition of the adult system." The authors' main concern is the extension of Labov's "variable rule" hypothesis (see Labov 1972 and elsewhere) to data from language acquisition, and therefore unfortunately the article contains no detailed data on the sequence of acquisition of SAI-related structures per se. However, the data that is supplied (and accompanying comments) make it clear that Jessie's acquisition of interrogative structures broadly parallels that of the Harvard children. First of all, the Labovs report the occurrence of formulaic WH-questions such as What's that ? at 27 months, prior to the use of auxiliaries in yes/no questions. This accords with the early use of routinized WH-forms in the language of the other children we have investigated. Second, at this same early stage Jessie went through a period where she produced WH-questions without auxiliaries, such as Where the boy ?; this resembles period A in S.'s development. Third, Jessie showed only sporadic use of questions in the early stages of her acquisition of interrogative structures, followed by a sudden leap in

frequency; again, this corresponds to the pattern we have observed in S.'s developing grammar (compare period A in (225) above with period B.) Fourth, there were no occurrences of WH-in-situ questions such as *He can't do it why ?, a generalization which appears to be exceptionless in the language of children learning English. Fifth, the Labovs refer to a stage (when the child was 46 months old - no MLU is supplied) where all yes/no questions were produced correctly - i.e., with inverted auxiliaries - but WH-questions were of four types: those with no auxiliaries, those with contracted auxiliaries, those with uncontracted auxiliaries but no inversion, and those with uncontracted auxiliaries and inversion. Examples of each type from Jessie's corpus are reproduced in (40) below:

(40) WH-question types in Jessie's corpus at 46 months:

(a) No auxiliary:

But how them buy their tents ?

(b) Contracted auxiliary:

Where's the chickens ?
What's that ?

(c) Uncontracted auxiliary, no inversion:

What that means ?
Why you said to daddy you might be kidding ?

(d) Uncontracted auxiliary, inversion:

How do babies get inside the mommies ?
Why are we going down ?

Of course, these are exactly the same four types of WH-question

that we have found in our investigation of WH-questions in the other children we have looked at. However, Jessie's grammatical development differs from that of the Harvard children in two significant ways, which we shall now discuss.

Firstly, Jessie's proportion of inverted to non-inverted WH-questions changed gradually over the course of the period in which she was studied from a low of 10% at 40 months to a high of 78% at 55 months, without any sudden transition that might indicate a "categorical" shift in grammatical function (note that the Labovs did not count WH-questions without auxiliaries separately; it might be that a more categorical picture would have emerged had they done so.)

Second, Jessie appeared to learn inversion separately for each WH-word; in fact, the bulk of the Labov and Labov paper is devoted to this issue. The Labovs report a "regular stratification of the four forms in the order how, where, what, why..."; why seems to discourage SAI to the extent that a special rule is needed of the form "Invert after all WH- words except why." The question of why such a stratification should exist is examined in detail by the Labovs, using Cedergren and Sankoff's (1974) method of multi-variate analysis. They distinguish eight sets of variables: (1) WH-form; (2) nature of subject NP (nominal versus pronominal, singular versus plural); (3) auxiliary form; (4) nature of complement (sentential versus non-sentential); (5) presence of contraction; (6) tense and aspect (past versus non-past, progressive versus non-progressive); (7) presence of negation; (8) effect of meta-

linguistic context (i.e., language games).

Such an analysis is potentially illuminating; unfortunately, however, we are only provided with a "preliminary" analysis over the whole of Jessie's corpus, which gives us no insights whatsoever into the development of her grammatical system. Perhaps this is a case of not being able to see the methodological wood for the trees. Nevertheless, the authors do provide us with a more fine-grained analysis of factors (1) and (5), the influence of a particular WH-word on the likelihood of inversion, and the effect of "contractability". Let us examine these in turn.

As far as factor (1) is concerned, the Labovs found a strong "lexical" effect on acquisition of SAI: examining the probability of inversion with individual WH-words at two month intervals, and factoring out the other variables detailed above, they found a consistent ranking of WH-words over an eight month period from about 45 months to about 53 months, in the following order: how, where, what, when, why. We will leave the question of why this ranking should exist until later in this chapter.

The other major findings in the Labov study concern the issue of "contractability" (factor 5). As we have seen, young children frequently misanalyze sequences consisting of a WH-phrase + contracted auxiliary; it might be expected, then, that WH-questions with contracted auxiliaries will have an unusually high proportion of "inverted" auxiliaries early in the development of interrogative structures (since such structures

will actually be equivalent to WH-questions without auxiliaries) and that the discrepancy will decrease as the child reanalyzes these sequences as "true" cases of SAI. This is in fact exactly what Labov and Labov report: in Jessie's corpus, the probability of "inversion" with contracted auxiliaries actually decreases from around $p_i = 0.75$ in a period from the age of 40-44 months to p_i = approximately 0.40 at 44-48 months. Moreover, this decrease in the probability of "inversion" in WH-questions with contracted auxiliaries coincides with a sudden rise in overall frequency of WH-questions during the same period; this lends support to the idea that an early, relatively primitive system for constructing WH-questions is being supplanted by a more sophisticated and general system at the stage in question. It should also be mentioned that the proportion of contracted auxiliaries with what and where (the two most frequent WH-phrases in the corpus) is much higher than with other WH-words; this probably reflects both the preponderance of what and where questions at early stages of syntactic acquisition (since they are conceptually "easier") and the higher proportion of contracted auxiliaries with these two WH-phrases in the adult system.

To conclude our discussion of the Labov paper, let us summarize the main points that have emerged:

(a) Jessie's development broadly parallels the sequence discovered by Bellugi; in particular, WH-questions appear

to follow a different developmental sequence from yes/no questions, with an initial preponderance of formulaic expressions, replaced at first by WH-questions with non-inverted auxiliaries, and then by adult-like use of SAI.

(b) On the other hand, there is little evidence of distinct "stages" in Jessie's acquisition of the auxiliary system (though it might be that a more fine-grained analysis would reveal them.) Rather, development towards the adult system seems to occur gradually.

(c) There is a strong lexical effect in the acquisition of SAI, with individual words varying widely in their ability to induce inversion at some points in the developmental continuum.

(d) At first, contracted auxiliaries are treated as unanalyzed lexical items, tending to artificially inflate the proportion of "inverted" <u>what</u> and <u>where</u> questions. As the child reanalyzes these sequences as genuine cases of SAI, this proportion decreases.

We have now examined four different sets of longtitudinal data in our investigation of the emergence of interrogative syntax, and come to four rather different sets of conclusions. At this point, we turn to cross-sectional studies; as we shall see, the perspective provided by large-scale sampling, far from

providing a resolution, seems at first sight to only add to the confusion.

5.4 The Acquisition of SAI: Cross-Sectional Studies

5.40 Introduction

In this section, we will be examining four cross-sectional studies on the acquisition of the interrogative system in English. Each employs a different methodology: Ingram and Tyack (1979) is based on spontaneous samples; Kuczaj and Brannick (1979) relies mainly on elicited imitation, with a supplementary grammaticality-judgement test; Erreich (1984) employs an elicited production technique; and Stromswold and Pinker (1986) is based mainly on the results of a judgement test. Moreover, the studies have rather different aims: while Ingram and Tyack (1979) and Erreich (1984) explicitly set out to assess the validity of Bellugi's SAI-acquisition stages, Kuczaj and Brannick (1979) is mainly concerned with undergeneralization, overgeneralization, and "lexical effects" in the acquisition of inversion in WH-questions, and the most significant results of the Stromswold and Pinker study concern inversion in non-matrix clauses. Finally, the subjects in the four studies are of different ages and at different points in the development of SAI.

Now, all this makes for a great deal of difficulty in comparing the results of the four studies. Nonetheless, they

provide a great deal of interesting data, which, together with the longtitudinal studies we have already investigated, can be pieced together to provide a quite substantial and detailed picture of SAI acquisition.

5.41 Ingram and Tyack (1979)

The first study we will examine is that of Ingram and Tyack (1979), which was designed explicitly to test the validity of the conclusions originally drawn by Bellugi. The methodology for this study consisted of the collection (by parents) of a "mini-corpus" of 225 questions for each of 21 children, matched by age and upper bound of utterance (UBU) to Bellugi's periods A-D. (41) (=Ingram and Tyack's Table II) shows the distribution of subjects across Bellugi's periods, age and MLU:

(41)

Period	Upper bound	Ranges	N	Age
A	7	4-7	7	2;4
B	9	8-9	3	2;9
C	11	10-11	5	2;10
D	13	12-13	4	3;4
E	14+	14-18	2	3;9

In analyzing the data, first of all, overall percentage of occurrence of auxiliaries in obligatory contexts was calculated for each UBU group. Then repetitions, subject questions, and questions without auxiliaries were omitted, and percentages of inversion in yes/no and WH-questions were calculated for each child.

The results of the study fail once again to support the universal existence of a stage such as Bellugi's period C. None of the 21 children inverted in yes/no questions but not in WH-questions. Indeed, the only discrepancy between the two types of question occurred with two (out of seven) children in the UBU level I group: these two had the opposite pattern, with WH-questions more likely to trigger inversion than yes/no questions. However, this "reverse" pattern is of dubious validity, since the criteria for "countability" failed to take into account the influence of WH-questions with contracted unanalyzed auxiliaries (which, as we have seen, are a notable feature of children's grammars at this stage). It seems likely that the children in question produced a large number of such questions, skewing the proportion of inversion in WH-questions, just as in the Labov and Labov study. In fact, Ingram and Tyack themselves make exactly the same point.

In fact, the authors report that only two of their subjects produced any uninverted WH-questions at all. Both of these children fell into UBU level II, and neither produced a high percentage of such questions (4 and 6 examples, compared to 19 and 43 with inversion, respectively). This pattern appears to resemble that of S. at period B - see (12) above - but unfortunately, since no figures are given for the number of WH-questions with contracted auxiliaries or without auxiliaries, we are unable to push the comparison further.

Another point made by the authors concerns the correlation between MLU (or in this case, the equivalent measure of UBU)

and acquisition of auxiliaries. Contrary to the findings of Bellugi, Ingram and Tyack report that the first auxiliaries occurred as early as period A, and then gradually increased in frequency until they reached Brown's (1973) criterion of 90% occurrence in obligatory contexts during Bellugi's period D (Brown's stage V). This should come as no surprise in the context of the current investigation; we have already seen in our examination of S.'s developing grammar that there is only a weak correlation between MLU/UBU and the development of the auxiliary system.

To summarize: the Ingram and Tyack study shows that a Bellugi-type period C is relatively infrequent across subjects, and certainly not an (overt) universal stage of syntactic acquisition. Unfortunately, the nature of the analysis employed (in particular, the lack of information on questions without auxiliaries or with contracted auxiliaries) prevents us from examining more closely the developmental patterns of the individual subjects; there is a tendency in cross-sectional studies of this kind to collapse data across subjects (perhaps understandable, in view of the time-consuming nature of the task of examining a large number of individual developmental sequences) and in doing so to obliterate the type of fine-grained analysis necessary for detailed investigation of emerging grammars.

5.42 Kuczaj and Brannick (1979)

Let us then turn to the second experimental study of the acquisition of interrogative structures, that of Kuczaj and Brannick (1979) - henceforth KB. In fact, their article contains three separate experiments, the first two involving an elicited imitation task and the third both elicited imitation and judgement tasks; since the third study is essentially methodological in nature - it concerns the interaction of grammaticality judgements with imitation abilities in older (6-year old) children, we will discuss only the first two studies here (see Chapter 1 for some comments concerning methodology).

The first task involved 20 children, divided into two groups according to age (average age of group I was 43 months, of group II 54 months). Each group was given 40 grammatical model questions containing modal auxiliaries to imitate, 5 each beginning with the WH-words who, what, where, why, how, when, how long and how come. Results were analyzed for correctness, number of omitted auxiliaries, number of misplaced auxiliaries, and lexical learning effects (number of children who scored 80% correct on a particular WH-word, and number of WH-words scored 80% correct/child).

Unsurprisingly, performance improved (significantly) over age; in particular, the number of ommitted auxiliaries decreased. More interestingly, the number of misplaced auxiliaries did not differ between the age groups; unfortunately, once again no more detailed analysis is possible, since results are collapsed across subjects in the two groups.

The most interesting results (and, indeed, the focus of the

study) concern lexical learning effects. KB found a fairly consistent ranking of WH-words across both age-groups and error patterns. (42) below shows these rankings (based on tables 2 and 3 in KB):

(42)a. Mean number of omitted modals (grouped across subjects and age, out of five):

Where	What	When	How	Who	Why	How come	How long
0.55	0.85	1.3	1.3	1.4	1.4	1.45	2.00

b. Mean number of misplaced modals (grouped across subjects and age, out of five):

What	Where	Why	When	How	Who	How long	How come
0.4	0.45	0.8	1.1	1.1	1.25	1.3	2.15

c. Mean number of total errors for Group I (grouped across subjects, out of five):

What	Where	When	How	Why	Who	How come	How long
0.7	0.7	1.45	1.5	1.55	1.6	1.8	2.00

d. Mean number of total errors for Group II (grouped across subjects, out of five):

Where	What	Why	How	When	Who	How long	How come
0.3	0.65	0.65	0.9	0.95	1.05	1.3	1.8

(42) shows that the ranking of WH-words is remarkably similar both across age and across error-type, with what and where consistently least likely to induce error, followed by how, when, why and who, followed by how long and how come. The only WH-word which seems to vary in a significant way is why, which shows a higher proportion of omission than misplacement errors

and a significant comparative improvement over age. These findings are also reflected in the number of children in each group who reached the 80% success criterion for a given WH-word: for group I, the ranking was: what/where, who, how/when, why, how long/how come; for group II it was Where, What, Why/how, how long, how come (see Table 4 in Kuczaj and Brannick). Note that a cross-sectional perspective throws doubt once again on a generalization reached on the basis of the study of one child; this time, however, it is the developmental pattern exemplified by the Labovs' child Jessie which proves to be exceptional, in that how, which consistently improved the probability of SAI in Jessie's grammar, has no such effect on KB's subjects - see (42b.) above. Of course, this does not cast doubt on the validity of the Labov's data, or on the validity of single subject studies; in fact, if anything, it shows that as far as the effect of individual WH-words on inversion is concerned, children may take rather different routes to the same adult system, an argument in favour of more fine-grained, single-subject analyses.

One other important aspect of KB's first study should be mentioned. This is the number of misplaced auxiliaries in how come questions (nearly twice as many as in the next most frequent misplacement-error inducing WH-question). As we have seen, how come does not induce inversion; therefore, such errors are cases of overgeneralization of SAI. This is particularly interesting in that a purely lexical account of the acquisition of SAI could not predict the existence of such

overgeneralization errors; thus they prove once and for all that children learning SAI are not engaged in a process of merely learning a series of syntactic frames by rote.

In fact, one of the purposes of KB.'s second experiment was to further investigate the learning of how come questions, and in particular their relation to why questions, which are semantically identical but syntactically distinct. For comparison, when and how long questions were also tested, since these are syntactically identical but semantically distinct.

Again, elicited imitation was used as the experimental method. This time, however, both grammatical and ungrammatical model sentences were used: 25 grammatical sentences consisting of 8 Why M(odal) NP VP ? questions, 4 When M NP VP ? questions, 4 How come NP M VP ? questions, 4 How come NP V+ed {NP,PP} ? questions, and 5 How long M NP VP ? questions; and 27 ungrammatical model sentences, consisting of 4 When NP M VP ? questions, 8 How come M NP VP ? questions, 4 Why NP M VP ? questions, 4 Why NP V+ed {NP,PP} ? questions, 4 How long NP M VP ? questions, and 3 semantically deviant questions in which how long is inappropriately substituted for when. Thus each child received a total of 52 model sentences.

Subjects were 60 children evenly divided into four age groups: Group I had a mean age of 43 months, Group II of 56 months, Group III of 65 months, and group IV of 80 months.

This time, KB looked at individual patterns of response. They found eleven different response patterns, as documented below in (43), which is based on KB's Table 12:

(43)

Response pattern	No. of subjects/group			
	I	II	III	IV
A. No use of modal auxiliaries	3	2	3	0
B. Misplace modal with why/when; omit modal with how come/ how long.	0	1	0	0
C. Misplace modal with why/when/ how long; omit modal with how come.	2	0	0	0
D. Use modal correctly with why; omit modal in other types.	1	1	0	0
E. Correct use of modal with why/ when; no use with how come/ how long.	4	2	1	0
F. Correct use of modal with why/ when; misplace modal with how long; no use with how come.	0	0	0	1
G. Correct use of modal in all types except how come, where no modal employed.	0	1	0	0
H. Correct use of modal in all types except how long, where no modal employed.	0	2	0	0
I. Correct use of modal with why/ when; misplace modal with how come; no use with how long.	2	2	2	0
J. Correct use of modal in all types except how come, where modal misplaced.	2	1	4	2
K. Correct use of modals in all types.	0	3	5	13

(Note: "correct" usage signifies 75% grammatical response. Note also that the table excludes the test sentences without modals.)

First of all, note that the types of response recorded by KB

fall into by now familiar patterns (with the exception of the overgeneralization of SAI in how come questions, to which we will return in a moment). These are of three types: questions with omitted auxiliaries; questions with "misplaced" (i.e., uninverted) auxiliaries; and adult structures. Now, leaving aside inversion with how come, if we abstract away from lexical effects and simply look at which children omit, misplace or correctly invert modal auxiliaries, the eleven response types fall into five broader categories, as illustrated below:

(44)

	Omit	Misplace	Invert	Pattern	No.
Type I	yes	no	no	A	8
Type II	yes	yes	no	B,C	3
Type III	yes	yes	yes	F	1
Type IV	yes	no	yes	D,E,G,H,I	18
Type V	no	no	yes	J,K	30

Significantly, two possible types are missing: the first would be equivalent to an "idealized" Bellugi period C, with misplacement but neither inversion nor omission; the second would involve inversion and misplacement, but no omission. Note also that only 4 children out of 60 make misplacement errors, confirming our suspicion (and that of Ingram and Tyack) that such errors are not a normal but an exceptional phenomenon in the development of the interrogative system.

Turning to the effect of the individual WH-phrases on the likelihood of inversion, it is obvious from a brief glance

at (43) that there is a consistent order of acquisition at work once again. The four WH-words employed in the study break down into two well-defined sets, the first containing why/when (set I), the other how long/how come (set II). The following table shows how these sets define the response patterns in (43):

(45)

Pattern	Omission	Misplacement	Correct
A	I,II	-	-
B	II	I	-
C	II	I,II	-
D	I,II	-	I
E	II	-	I
F,I	II	II	I
G,H	II	-	I,II
J	-	II	I,II
K	-	-	I,II

(Note: incorrect how come imitations are included).

(45) shows clearly that sets (I) and (II) are implicationally ordered along the developmental sequence, lending support to the order of acquisition found in KB's first experiment (see (42) above) and, incidentally, adding credibility to the sequence itself.

Next, let us turn to the issue of overgeneralization with how come. First of all, note that misplacement errors (i.e., inversion) with how come exceed misplacement errors (i.e.,non-

inversion) with the other WH-phrases by a factor of five-to-one, as shown in (46), where the total number of each type of response for each WH-word is shown:

(46)

	omission	misplacement	correct
why	8	3	49
when	10	3	47
how long	26	3	31
how come	22	15	23

Clearly, overgeneralization with how come is a more likely type of error than "undergeneralization" with other WH-phrases. It also appears, perhaps not surprisingly, at a later stage; this is illustrated by (47), which shows the number of children (out of 15) making each type of error for each age group:

(47)

Age group	omission	undergeneral.	overgeneral.
I	12	2	4
II	11	1	3
III	6	0	6
IV	1	1	2

While both omission and undergeneralization (non-inversion) errors decline with age, overgeneralization (inversion) errors with how come rise to a peak in group III (five-year-olds) with a subsequent decline. It thus appears clear that the latter type of error is a product of a more advanced stage in the development of SAI.

One other point important point concerning the how come overgeneralization errors emerges from KB's study 2. Recall that test questions with how come were given both with modals and with verbs in the simple past tense. As KB point out, if children who overgeneralize inversion to how come in the modal cases have a truly general rule of SAI, they should be expected to invert in the past tense cases as well (presumably via SAI and insertion of auxiliary do). However, this was not the case: out of the 15 children who overgeneralized SAI in the modal questions, only five inverted in the simple past tense questions. This raises a number of non-trivial issues concerning rule overgeneralization, which we will address in the next section.

Based on the KB studies, the following picture then emerges of developmental patterns in the acquisition of SAI. At first, WH-questions are characterized by a high rate of auxiliary omission, occasionally accompanied - at a slightly later but overlapping stage - by non-inversion. As the adult system develops, both these types of errors disappear, and overgeneralization errors with how come appear (though such errors are still comparatively infrequent). Finally, these errors disappear and the adult system is acquired. KB also found a consistent order of SAI-acquisition with WH-words: what and where come first, followed by why, when, how, and who, with auxiliary cooccurrence restrictions in how long and how come questions acquired last at each "stage" of acquisition.

5.43 Erreich (1984)

Bearing in mind these conclusions, let us then turn to Erreich (1984). Erreich employed an elicited production technique, designed to enhance spontaneous use of yes/no- and WH-questions by her subjects; elicitations were "strategically integrated" into free-play sessions involving experimenter, parent, and child. Subjects were 18 children ranging in age from 29 months to 36 months and from MLU stage III to MLU stage V (mean age was 32 months, mean MLU 3.33, i.e., stage IV). These children were selected from an original number of 41; the other 23 were excluded either because they exhibited adult-like inversion patterns (14), employed auxiliaries "well below 50% of the time" (6) or were uncooperative subjects (3).

Scoring was for "clear cases of inversion and clear cases of non-inversion": this entailed the omission of repetitions, subject-questions, questions without auxiliaries, and questions containing auxiliary overmarking errors (see Chapter 3). Scoring was by type rather than token, in order to avoid the possibility of routinized formulas such as What's that artificially inflating the frequency of a given question-type. A further subdivision involved what Erreich (misleadingly) calls verbal type, i.e., the number of different auxiliaries employed by a given child in formulating a given type of question. This was used as an index of productivity. Questions were then divided into four categories, according to question-type (WH versus yes/no) and whether inversion took place.

On the basis of results obtained over an average of 2.1 hours of transcript/child, Erreich distinguished five response patterns, based on a productivity criterion of at least two verbal types/form. These are detailed below in (48), based on Erreich's table 1:

(48)

Pattern	no.	Yes/no		WH	
		+INV	-INV	+INV	-INV
A	1	-	4/5	2/20	-
B	1	4/5	6/10	2/16	1/1
C	1	5/5	9/15	1/10	8/12
D	5	0.8/0.8	12.8/18.2	10/12	5/21.2
E	10	7.5/10.1	13.8/20.1	8.7/12.6	5.4/20.9

(The figure before the slash is the mean no. of verbal types/question type; that after the slash is the mean no. of total questions/question type).

The first thing to notice about (48) is that only two out of the five response patterns contain a significant number of children. These patterns (D and E) are classed by Erreich as the "non-inverted/optional" and "optional/optional" groups, respectively. In fact, patterns A and B can probably be discounted, since both contain only two verbal types (barely above criterion) in WH-questions; this suggests that these children have memorized a couple of WH-formulas rather than acquired a productive strategy for integrating the auxiliary system with the WH-system. Notice also that the "classical" Bellugi "inversion/non-inversion stage is represented by only

one child (pattern C above).

Now, both patterns C and D appear to have an "optional" rule of SAI in WH-questions; they differ in that pattern D has no SAI in yes/no questions, while pattern E has SAI applying optionally in yes/no questions. Let us deal with the yes/no and WH-cases separately, beginning with the latter.

A glance at (48) will reveal that the pattern of inversion in WH-questions is almost identical in D and E. Of major interest is the ratio of the number of verbal types to the total number of questions in the inverted and uninverted cases. In the latter, it is (1 : 1.31); in the former, (1 : 4.05). In fact, the ratio for uninverted WH-questions resembles the ratios for both inverted and univerted yes/no questions in patterns D and E , which are (1 : 1.31) and (1 : 1.46), respectively. In other words, the ratio for inverted WH-questions differs very significantly from all the others. Now, if we take these ratios as a more sensitive index of productivity than Erreich's criterion of "at least two verbal types", it becomes clear that inverted WH-questions are being used far less productively than any of the other question types. This allows us a rather more refined analysis of the "optional" nature of inversion in WH-questions for patterns E and D; what appears to be going on is that these children are using a large number of formulaic WH-routines, giving the impression that they are optionally applying SAI in WH-questions. While the nature of the data-presentation in

Erreich's paper prevent us from examining this claim in more detail, it should be noted that for pattern E (the only group large enough for statistical significance to be reached) Erreich found that there was a slight but significant difference between the overall ocurrence of auxiliaries in WH-questions and in yes/no questions (77% for WH-questions, 90% for yes/no questions); again, this indicates that "inversion" in WH-questions is of a rather different status than that in yes/no questions. Obviously, it would be interesting to check the percentage of contracted auxiliaries in the "inverted" WH-questions of patterns D and E, as well as the number of WH-questions containing omitted auxiliaries; unfortunately, Erreich does not distinguish between contracted and non-contracted auxiliaries, and, as noted above, questions with omitted auxiliaries were not included in her scoring system. Nevertheless, we conclude on the basis of the available evidence that "optional" inversion in WH-questions is an epiphenomenon; children who appear to apply SAI optionally are actuallly not employing it at all.

Given this perspective, patterns D and E fall in line - as far as WH-questions are concerned - with pattern C, giving unexpected <u>support</u> to Bellugi's original claims concerning the "non-inversion" period. Children in D and E are employing a "mixed" strategy, using formulaic "inverted" WH-questions as well as uninverted WH-questions; there is no evidence, however, that they are employing an adult SAI rule in WH-fronting structures.

This support must be qualified, however, by Erreich's data on yes/no questions; here, there is little doubt that Bellugi's claim concerning the relation between yes/no and WH-questions is false, or rather, much too limited. Thus, while (on our reanalysis of Erreich's data) pattern E children appear to fall under Bellugi's period C (inversion in yes/no but not in WH-questions), pattern D children, who do not invert in yes/no questions, do not.

As for the development of SAI in yes/no questions, two points are worth making: first, inversion is always optional, even in the adult grammar, and second, there is a wide range of individual variation in the actual proportion of inverted to uninverted yes/no questions recorded in the data. The first point explains why there are no children with an obligatory rule of SAI in yes/no questions: there is no reason to postulate such a rule. The second point illustrates the problematic status of the theoretical construct "optional"; thus, the children with "optional" yes/no inversion displayed the following range of "optionality" in their use of SAI in yes/no questions:

(49)

Child	% of inverted auxiliaries in yes/no questions
Lauren	33
Mark	25
Eddie	87
Leo	88
Jamie	56
Tommy	84
Joshua	55
Ben	60
Scott	51
Jana	33
Andrew	65
Candice	50

It appears likely that the variation observed above is a result of "cognitive style" and other variable factors, rather than being linguistically determined; as noted by Erreich, inversion in yes/no questions is necessary neither for linguistic nor for communicative reasons (young children as well as adults can question using intonation only). The status of WH-inversion and yes/no inversion is thus rather different: the latter, indeed, might even be classified as a "stylistic" rule.

Several other aspects of Erreich's study deserve mention. First of all, the author found that negation had a strongly inhibitory effect on inversion, confirming earlier reports to that effect. However, pace Bellugi, there was no difference in this respect between WH- and yes/no- questions: none of the 16 negative questions in the data-base (5 yes/no, 11 WH) showed inversion.

Second, analyzing the effect of individual WH-words on the likelihood of inversion, Erreich found the order illustrated in (50)(= Erreich's Table 4):

(50)

WH-term	Non-inverted		Inverted	
	N	%	N	%
What	64	24	199	76
Where	43	28	112	72
Why	51	77	15	23
When	23	92	2	8
How	13	72	5	28
Who	1	6	15	94
Other	4	33	8	67
Which (one)	0	0	8	100

Thus, we get the following ranking of WH-phrases most likely to

induce inversion: which one and who, followed by what, where, and "other", followed by how, why, and finally when. Compare this order with that found by Kuczaj and Brannick and reported in (46) above.

Finally, Erreich reports that only four errors were found involving inversion in embedded WH-questions: this accords with previous work, but not with the findings of Stromswold and Pinker (1986), to which we now turn.

5.44 Stromswold and Pinker (1986)

Unfortunately, I have no direct access at present to the Stromswold and Pinker study; accordingly, our review of their work will inevitably be cursory. This is a pity, because the preliminary report in Stromswold and Pinker (1986) contains some interesting and unexpected findings.

Perhaps the most interesting aspect of the study is that, contrary to previous findings, the authors report - from a survey of WH-questions in ten children's spontaneous corpora - that inversion errors in embedded questions are quite frequent. They back this up with a judgement test involving 22 preschoolers, who judged root questions without inversion (e.g., Who Ernie is ?) significantly worse than inversions in embedded contexts (e.g., I know who is Ernie).

Other findings reported from the survey included the following: (i) performance varied for individual WH-words; (ii) subject questions apeared before object questions; (iii) only

four errors were discovered involving inversion with a non-auxiliary verb; (iv) there were no examples of questions involving WH-in-situ. All of these findings confirm conclusions we have already reached on the basis of our investigation so far.

5.45 Summary

In this section, we will atempt to summarize the confusing array of data which we have been examining. We will deal with the following key questions:

(a) The proportion of children who pass through a "non-inverting" stage.
(b) The relation between WH-questions and yes-no questions in the non-inverting stage.
(c) "Lexical" effects involving WH-phrases.
(d) "Overgeneralizations" involving <u>how come</u>.
(e) The effect of negation on SAI.
(f) SAI in embedded clauses.

We will examine these questions one by one.

(a) First of all, we turn to one of the most contentious issues in the study of SAI acquisition: the question of how many children actually do pass through the "non-inverting" stage. It is clear that the stage in question is not a universal component

of the development of SAI - our analysis of S.'s grammar, for example, shows no period of non-inversion - but it is unclear exactly what percentage of children learning English do go through a period of development marked by consistent non-inversion. Moreover, the cross-sectional studies we have examined seem to disagree on this question. (51) below shows the percentage of children in each study who made non-inversion errors:

(51)

	N	U	%
Ingram and Tyack	21	2	9.5
Kuczaj and Brannick	60	4	6.5
Erreich	43	17	39.5

KEY: N = number of subjects
U = number of non-inverting subjects
% = percentage of non-inverting subjects

While Ingram and Tyack and Kuczaj and Brannick report very few subjects who produce non-inverted questions, the figure for Erreich's study is much higher, even when the subjects she rejected are taken into account.

The chief reason for this discrepancy appears to be the different sampling techniques used in the three studies. Ingram and Tyack chose subjects on the criterion of UBU (MLU) stages, deliberately selecting children at each of stages I-V. Kuczaj and Brannick selected subjects on the basis of age, dividing the sample into groups of 3, 4, 5, and 6 year old children. Erreich employed both MLU and age, selecting only children in

the 1;10 to 3;3 age-range, with MLU scores ranging from 2.66 (i.e, stage III) to 4.26 (i.e., stage V). Thus whereas the Ingram and Tyack and Kuczaj and Brannick studies were designed to test children over a broad age/MLU range, Erreich's study was specifically targeted at the age and MLU value where children are most likely to make non-inversion errors.

Note also that the youngest children in the Kuczaj and Brannick study were 3;0, whereas the average age of the 18 children ultimately selected by Erreich was 2;8. In other words, these two studies were investigating two distinct periods of SAI development - and indeed, this is reflected in their results.

Now, suppose we narrow down the number of "potential non-inverters" in the Ingram and Tyack and Kuczaj and Brannick studies by eliminating subjects with MLU below 2.5 (i.e, stage II) or above 4.5 (i.e, late stage V) and age below 2;0 or above 4;0. In that case, the percentage of non-inverters rises considerably, as can be seen from (52):

(51)

	N	U	%
Ingram and Tyack	12	2	16.5
Kuczaj and Brannick	13	2	15.5
Erreich	43	17	39.5

KEY: N = number of subjects
U = number of non-inverting subjects
% = percentage of non-inverting subjects

However, a discrepancy of more than 2:1 still exists. For the

Kuczaj and Brannick study, this can be explained fairly simply in terms of age: their 3;0 starting point is more or less in the middle of the period when children are most likely to produce non-inversion errors, and thus they are only likely to catch about 50% of the children who pass through a non-inversion stage. The difference between Erreich's findings and those of Ingram and Tyack, on the other hand, seems to derive from their different scoring systems. Erreich scored a child as an "optional" non-inverter on the basis of non-inversion in questions containing at least two different verbal types, whereas Ingram and Tyack rely on percentage of inversion. Now, suppose we apply Erreich's scoring system to Ingram and Tyack's subjects (with a bit of educated guesswork; since Ingram and Tyack do not give figures for verbal types, we have translated "at least two verbal types" into "at least five questions"). In that case, at least two more children qualify as non-inverters, bringing the percentage up to around 30% of potential non-inverters.

We conclude that children who pass through some kind of "non-inverting stage" constitute around 25-35% of all those learning English as a first language. This makes sense both of Bellugi's early claims (since it is not statistically improbable that all three Harvard children should have been non-inverters - and even then, note that Adam was the only one of the three who showed the pattern "clearly") and at the same time explains subsequent findings that non-inversion is by no means universal, and in fact actually seems to characterize only a minority of

children's developing grammars.

(b) Let us then turn to another vexed question in the analysis of children's SAI development: the relationship of SAI in yes-no questions to SAI in WH-questions during the "non-inverting stage". Once again, we have seen that Bellugi's original claim (that inversion in yes-no questions always precedes that in WH-questions) is false; K., for example, was less likely to apply SAI in yes/no questions than in WH-questions. And once again, we are faced with the question of how general the Bellugi pattern in fact is.

Suppose, as a null hypothesis, we were to claim that inversion in yes-no and WH-questions was learnt quite separately. In that case, we would expect any one of the following three patterns during "the non-inversion stage": (i) inversion in yes-no but not in WH; (ii) inversion in WH but not in yes-no; and (iii) inversion in neither. The issue is of course considerably confused by the fact that inversion in yes-no questions is optional in the adult grammar, and by the influence of (WH + contracted auxiliary) sequences, which tend to inflate the number of "inverted" WH-questions. Nonetheless, there does appear to be a significant gap in this paradigm: of the four studies that have relevant data on the issue, Bellugi's findings support a developmental pattern such as (i), our study of K. seems best characterized by pattern (iii), Ingram and Tyack's two non-inverters also follow pattern (iii), and the two major non-inverting groups in Erreich's study, the

"non-inverted/optional" and "optional/optional" fall under patterns (iii) and (i), respectively; but none of the subjects in these studies show a "non-inverted/inverted" distribution corresponding to pattern (ii).

Once again, then, we have found Bellugi's claims to be half-supported by the data; while there are children who follow the developmental schema exemplified by her subjects, there are also a significant number who invert in neither type of question. However, the development of inversion in WH- and yes-no questions is not entirely independent, since there do not appear to be any children who pass through a stage involving inversion in WH-questions only. We will return to the significance of these findings in 5.7 below.

(c) Next, we turn to the issue of "lexical specificity" in the acquisition of SAI in WH-questions. While a number of the studies we have examined have pointed out that SAI is acquired WH-word by WH-word, I think it is significant that they all propose <u>different</u> orders for this developmental sequence. (52) below summarizes those proposed in the studies we have reviewed:

(52)

Labov & Labov:	how, where, what, when, why.
Kuczaj & Brannick:	where, what, who, how, why, when, how long, how come.
Erreich:	which one, who, what, where, how, why, when.
K.:	how X, where, why, what, who, what X.

(Note: the order for K. is based on percentage inversion over the whole six month period of investigation, and excludes contracted auxiliaries).

In other words, while children do appear to acquire SAI WH-word by WH-word, there is much individual variation as to which WH-words are more likely to induce inversion. Moreover, the probability of a given WH-word inducing inversion changes over time even in the grammar of a single child; in (53) below, we have ranked WH-words according to their inversion-inducing probability for K.'s Periods II, IV, and V (I and III are omitted since there were no inverted auxiliaries during those periods):

(53) Period II: why, what/where
Period IV: what, why/where
Period V : how X, who, why/what, how/where/what X

Thus the probability of a given WH-word inducing inversion is subject to intra-subject variability as well as inter-subject variability. Moreover, it is subject to token- as well as type-variability; the same WH-word may or may not induce inversion during the same period for the same child. (54) shows three <u>what</u> questions produced spontaneously by K. during the same session, when he was 35 months old and his MLU was 3.56:

(54)a. What is this ?

b. What's that ?

c. What's this is ?

So we see that variability in WH-induced SAI is not simply an inter-subject, type-variable "lexical learning" effect, but a more fundamental characteristic of SAI development, involving intra-subject and token-variability. We will return to the significance of this finding in 5.7 below.

(d) The only study to deal with the "overgeneralization" of how come is that of Kuczaj and Brannick. This is almost certainly due to the fact that this study dealt with a later period of acquisition than any of the others we have reviewed. In fact, errors with how come reach a peak around the age of five years, as shown in (47) above; this is well after the non-inversion stage, at a period when in general children have acquired an adult-like rule of SAI. This in turn indicates that SAI acquisition in WH-questions shows a "U-shaped learning curve" of the type we have associated with closed-class learning: initial undergeneralization is followed by limited overegularization of irregular forms and finally acquisition of the adult system. Once again, we will explore the significance of this finding in 5.7 below.

(e) The inhibitory effect of negation on inversion, first reported by Bellugi, seems to be supported by the only other study to address the issue, that of Erreich (1984). However,

while Bellugi claimed that negation only affected the probability of inversion in WH-questions, Erreich reports that negation inhibited negation in both yes-no and WH-questions. Unfortunately, children spontaneously produce very few negative questions; clearly, additional experimental studies would be very helpful in resolving this question.

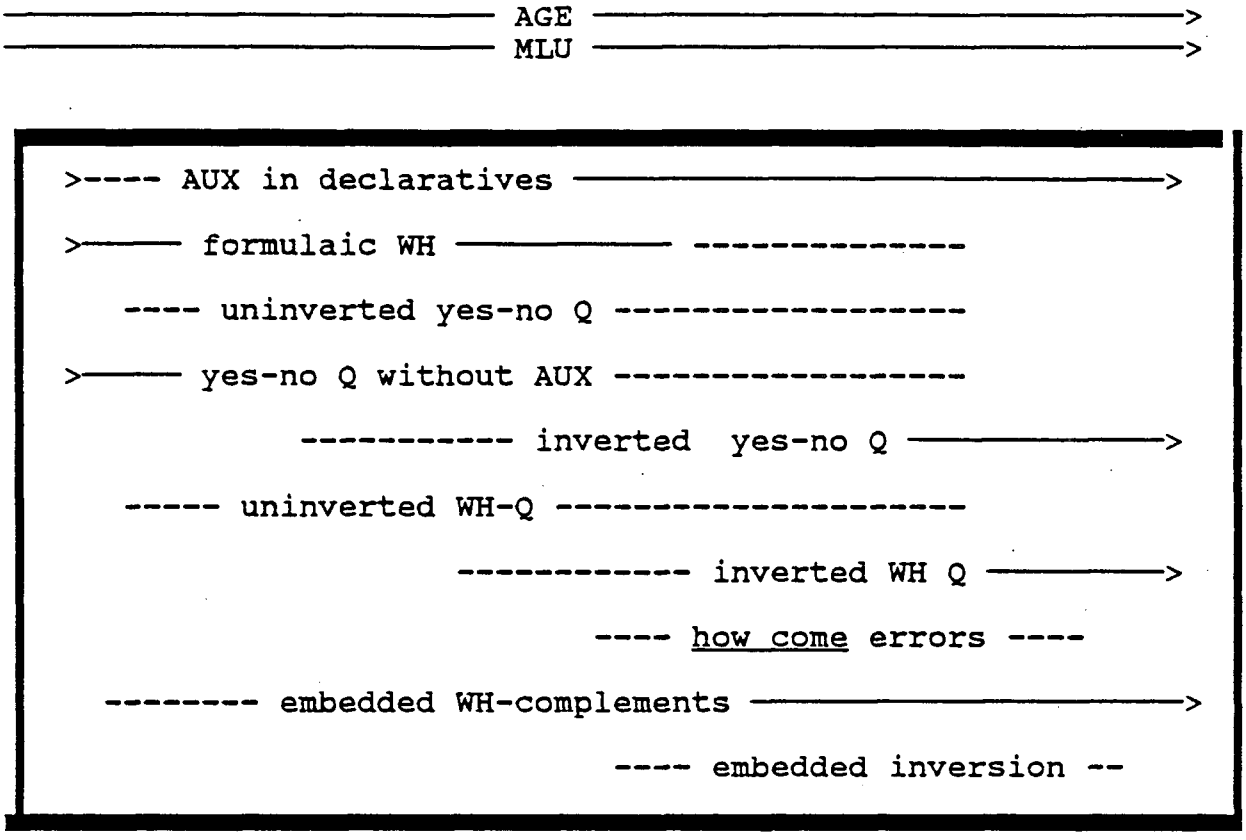
(f) Finally, we turn to the question of inversion in embedded clauses. Once again, there appears to be disagreement in the data here. On the one hand, Bellugi reports very few errors involving inversion in subordinate clauses (Adam only produced one, according to Pinker 1984), K. only produced one inverted auxiliary in an embedded environment, S. produced none, and Erreich reports that only two out of her 18 subjects produced such errors (two each). On the other hand, Stromswold and Pinker (1986) report that inversion errors in embedded clauses occur much more frequently than was previously thought; moreover, they provide experimental evidence to back this claim.

Once again, the age of the subjects who make errors with inversion in embedded clauses is likely to be a crucial factor in resolving this apparent contradiction. It appears highly probable that this type of error, like over-application of SAI in <u>how come</u> questions, occurs at a comparatively late stage in the development of SAI, and thus constitutes a case of limited overgeneralization. Unfortunately, I have no access to information concerning the precise age of the "pre-schoolers"

in the Pinker and Stronmswold study; however, if, as I suspect, children who overgeneralize SAI to subordinate clauses turn out to be in the same age-range as those who invert in how come questions, such errors will provide further support for the idea that the acquisition of SAI involves closed-class learning, with an attendent "U-shaped" learning curve.

On the basis of the preceding discussion, we can now produce a tentative picture of the acquisition of SAI in English. Rather than proposing a fixed set of stages, however, we will represent the developmental sequence in the form of the flow-chart in (55):

(55)



(Note: a dotted line indicates that not every child produces the structure in question).

(55) summarizes most of the main points of the preceding discussion.

We will now turn to a brief review of the acquisition of INFL2 structures, which will provide an interesting contrast to the acquisition of SAI.

5.5 The Acquisition of INFL2

5.50 Introduction

In this section, we will review what is known about the

acquisition of INFL2 - principally in German, but with some additional data from Dutch. My main source is Clahsen and Smolka (1986), which summarizes the results of some eighteen studies of the acquisition of German syntax. Dutch data is from Kaper (1984). Clahsen and Smolka divide the development of INFL2 into four stages, based on an implicational scaling technique for assessing syntactic development; we will deal with the first two in 5.51, and the latter two in 5.52. There appears to be a remarkable unanimity concerning the validity of these stages in the INFL2 acquisition literature - perhaps all the more surprising in view of the confusion and disagreement which characterizes work on the acquisition of SAI. In fact, as we shall argue in 5.7, this is not accidental, but follows from the very different types of learning involved in acquiring these two closely related syntactic processes.

5.51 Stages I and II

Since in German and Dutch INFL2 applies in almost all main clauses, we might expect children learning these languages to start out by producing sentences with XP V XP order. Such sentences are attested early in the syntactic development of German- and Dutch-speaking children; however, the predominant type of word-order at this stage is <u>verb-final</u>. Examples are given below for Clahsen and Smolka's Stage I.

(56)a. (G) boden bürs M (29 mos., 1.63)

floor brush

b. (G) purzel pierkorb rausräum — M (29 mos., 1.63)
P. paper basket remove

c. (G) Julia schieben zuch — M (27 mos., 1.54)
J. push train

d. (D) Nel stoel halen - gang — H (23 mos.)
N. chair carry corridor

e. (D) Deze boek hebben, Mamma — H (27 mos.)
These book have M.

f. (D) Ik heeft het, Pappa — H (29 mos.)
I has it P.

Note (G) = German, (D) = Dutch. Age and MLU are given where known.

Now, why should children who normally only hear verb-final order in subordinate clauses nonetheless produce a majority of verb-final utterances in matrix clauses ? There are only two possible viable explanations. Firstly, it could be that children selectively attend to embedded clauses in the input. As we shall argue more thoroughly in 5.64, this is both theoretically and empirically implausible. The only alternative, however, is to assume that even very young children are performing quite sophisticated pre-productive syntactic analysis, abstracting away from INFL2 effects to acquire the basic (SOV) word order of German and Dutch. Thus, in a sense, such children must already know how INFL2 works in order to acquire it. We will return to this somewhat paradoxical conclusion in 5.7 below.

It might be asked exactly what kind of evidence the Stage I learner could employ in order to arrive at an SOV analysis for

German/Dutch. Even assuming that subordinate clauses do not in general count as relevant evidence (i.e, restricting the PLD to degree 0), there are in fact at least three relevant kinds of evidence. The most obvious of all is the fact that the S-initial ("topic") position in INFL2 clauses does not bear a particular θ-role (and in fact need not even be an NP); this will tell the child that some type of topicalization must be going on. Now, assuming that the child has access to principles of UG, obligatory topicalization of the type exemplified by INFL2 languages will automatically entail [I,V] movement, in order to link the A'-type P-subject to AGR and thus satisfy conditions on P-licensing (see Chapter IV for details). However, such evidence will not tell the child the base structure from which the topicalization/[I,V] movement structure is derived; in order to determine the latter, evidence must be available to link the fronted [I,V] constituent to a particular base-position. Such evidence is provided by two types of structure: those containing an auxiliary as well as a main verb, where the finite auxiliary will front, leaving the verb in final position; and those containing verb-particle sequences, where the particle is "stranded" in V-final position when the verb fronts.

Thus there <u>is</u> degree-0 evidence for a basic SOV order in German and Dutch; however, the child's grammar must be sophisticated enough to use it. In particular, note that we assume the child has access to the AGR-linking condition, even though at this stage in development there is no indication that

children learning German and Dutch have grasped the morphological agreement system; once again, we find ourselves in the paradoxical position of attributing to the child a large amount of syntactic knowledge which s/he appears to be able to employ pre-productively but not in production.

Let us then turn to Stage II, which is distinguished from Stage I chiefly by the development of V-particle constructions. These appear uniquely in verb-final positions, as shown by the following examples:

(57)a.	(G)	da pappa anrufen there P. up-call	M (32 mos., 1.87)
b.	(G)	die auto hier boot umkipp the car here boat upset	M (33 mos., 1.72)
c.	(D)	kop thee opdronken cup tea up-drunk	E (33 mos.)
d.	(D)	kussen opstaan cushion on-stand	E (30 mos.)

These examples are interesting because, as indicated above, discontinuous verb-particle constructions probably constitute crucial evidence for verb-final order, yet the child never produces such constructions. Once again, we are presented with a curious contrast between analytical sophistication and an apparent inability to apply this sophistication to syntactic production.

5.52 Stages III and IV

The most striking developmental change in the acquisition of INFL2 structures occurs in the transition from Stage II to Stage III. At Stage III, all finite verbs occur in INFL2 or INFL1 position, and non-finite verbs appear in final position:

(58)a. (G) ich hab ihn aufsetzt — M (37 mos., 3.58)
I have it on-put

b. (G) jetzt hast du sechs — M (37 mos., 3.58)
now have you six

c. (G) die eisenbahn hat die Julia puttmach — M (36 mos., 3.16)
the railway has J. destroyed

d. (D) twee jammen hebben we — H (38 mos.)
two jams have we

e. (D) gisteren had Erik die omgegooid — H (38 mos.)
yesterday had E. that upset

f. (D) hier heeft ik gevallen — H (38 mos.)
here have I fallen

Clahsen (1982) reports that the transition between Stage II and Stage III occurs very rapidly; in the course of a single month, the proportion of verb-second structures used by the three children in his longtitudinal study rose from 20% to 40% (Stages I and II) to 90% (Stage III); this pattern is reported by all the other studies reviewed in Clahsen and Smolka (1986). This rapid acquisition sequence, of course, contrasts with the slow and uneven development characteristic of the aquisition of SAI in English.

What triggers the change ? According to Clahsen (1984), the transition from Stage II to III is correlated in German with the acquisition of subject-verb agreement morphology. At Stage I, his subjects only employed the infinitival form -n, the 0 base-form and the "dominant" third-person suffix -t; this is typical of the early acquisition of inflectional systems, where one salient ending is often overgeneralized throughout a paradigm (see Slobin 1985 for a discussion of "inflectional imperialism"). Stage II is a "transitional" phase, where the first-person suffix -e begins to appear; however, it is not until Stage III, with the acquisition of the second-person ending -st that the entire paradigm is learnt (see Clahsen 1984 for details). As soon as the child has mastered the paradigm, the transition from Stage II to III takes place.

It thus appears that the acquisition of INFL2 is dependent on the acquisition of agreement morphology. This is a finding of some significance for our account of SAI-type rules, which, of course, is crucially based on the concepts of AGR-linking and AGR-domain (once again, see Chapter 4 for details). Note, however, that L1 learners of German do not suddenly "realize" that INFL2 is tied to AGR (and hence, to the finiteness of the verb) once they have productive control of the appropriate verb morphology; as we have seen, children learning German grasp the AGR-related nature of the INFL2 process at a much earlier age (in Stage I). Rather, it appears that their productive control of INFL2 is dependent on the acquisition of specific subject-verb agreement inflections; once again, we see what Kuczaj and

Maratsos (1983) refer to as "a peculiar mixture of general and specific knowledge" in the acquisition of syntax.

Stage IV marks the acquisition of subordinate clauses, which Clahsen and Smolka report are always produced with the (correct) verb-final order:

(59)a. (G) ich will mal sehen ob das M (42 mos., 4.53)
I want just see if that

schwarz ist
black is

b. (G) weil ich ein stern machen M (41 mos., 4.07)
because I a star make

muβ
must

c. (D) Pappa, hoe lang duurt dat H (46 mos.)
P. how long takes that

Mamma klaar is ?
M. ready is

d. (D) Mamma, weet je wat vlug H (46 mos.)
M. know you what fast

lopen is ?
walking is

In contrast to children learning SAI in English, it appears that INFL2 learners seldom if ever make embedded inversion errors. Clahsen and Smolka report that they found no errors of this type in their review of the literature on the acquisition of German.

5.53 Summary

Even from this brief review of the development of INFL2 structures, it is obvious that there are significant differences between INFL2 acquisition and the acquisition of SAI. We will return to the theoretical significance of these differences in 5.7. Before going on to examine various attempts at explaining SAI, however, let us summarize the most salient aspects of the acquisition of INFL2 structures:

(a) Children learning verb-final INFL2 languages initially produce most clauses with verb-final order, even though it is unlikely that a majority of clauses in the input have this order.

(b) The transition from verb-final to INFL2 order takes place rapidly, in contrast to the prolonged, variable development characteristic of the acquisition of SAI.

(c) The transition from verb-final to INFL2 order seems contingent on the acquisition of the subject-verb agreement paradigm in German.

(d) When they begin to produce subordinate clauses, German- and Dutch-speaking children invariably employ verb-final order, and do not overgeneralize INFL2 to embedded environments.

5.6 Previous Accounts of the Acquisition of SAI

5.60 Introduction

In this section we will review previous attempts at accounting for the peculiar developmental sequences which characterize the acquisition of SAI, notably those of Bellugi (1971), Klein (1982), and Pinker (1984). There have been three main approaches taken towards the problem of explaining SAI development: those based on properties of intermediate grammars (including both Klein's and Pinker's accounts); those based on the influence of performance factors (most notably, that of Bellugi); and those based on features of the input (including the explanations put forward by Brown and Hanlon 1970 and Maratsos 1978). We shall argue that no single one of these approaches can adequately handle the complex array of data which, as we have seen, characterizes the acquisition of SAI; in 5.7 we will suggest an alternative based on the interaction of several different factors, which will lead us in 5.8 to propose a model for syntactic learning which differs in several significant ways from the linear-parametric model which we originally set out to investigate.

5.61 Bellugi's Account of SAI Development

Let us then begin with the oldest and best-known explanation for SAI development, that of Ursula Bellugi (see Klima and Bellugi 1966, Bellugi 1967, Bellugi 1971). While based on data from the Harvard study which, as we have seen, is only

partially representative, Bellugi's account is interesting in that it represents a particular type of approach to explanation in language acquisition, which appeals to developmental discrepancies in performance factors as a means of accounting for what appear to be grammatical differences between adults and children (see Chapter 1 for some general comments on the relation of competence to performance factors in language acquisition). In fact, we have already come across this type of explanation in Chapter 2, where we discussed "sentence-span limitations" as a means of accounting for the "missing subject stage" in the acquisition of English; there, we argued that such accounts lack explanatory potential as long as the performance factors in question remain unelucidated - as is certainly the case with sentence-span constraints.

In other words, the success of a performance-based explanation depends on the success of its treatment of the performance factors involved. With this in mind, let us then proceed to examine Bellugi's explanation for the non-inversion stage in the acquisition of SAI.

First of all, recall Bellugi's empirical claims: inversion in yes-no questions precedes inversion in WH-questions, and inversion in affirmative WH-questions precedes inversion in negative WH-questions. We have seen that these claims are only partially borne out by a wider sample of subjects; nevertheless, for argument's sake, let us assume for the moment that they are correct.

Now, the central component of Bellugi's explanation for this

developmental sequence is the notion of "cumulative complexity"; SAI applies in yes-no questions before it applies in WH-questions because the latter are more "complex" than the former. Likewise, negative WH-questions are more "complex" than affirmative WH-questions, therefore they are acquired later.

Several questions immediately arise as to the nature of this "complexity limitation". In particular, is it to be defined syntactically or with reference to some other cognitive system ? And if the latter option is chosen, how is the notion of complexity to be defined so that it will have the right influence on syntactic acquisition ?

In fact, Bellugi herself is rather vague on these crucial questions, and not without reason; her explanation was first conceived in the heyday of the Derivational Theory of Complexity (DTC - see Chapter 1 for discussion), when it was thought that there was a relatively simple and transparent relation between rules of grammar and their realization in various types of linguistic performance, amongst which, of course, we can include real-time language acquisition. Now, given an acquisition theory which incorporates some version of the DTC, the notion of "cumulative complexity" is easy to interpret; essentially, we can derive an index of syntactic complexity simply by adding up the number of transformations that have applied to a given structure. Thus a negative WH-question, which is derived via the application of three major transformations (SAI, WH-movement, and not-placement) is more complex than an affirmative question, which only involves the

first two, which is in turn more complex than a yes-no question, where only SAI has applied. A child with a "complexity limitation" of 1 will then only be able to apply inversion in yes-no questions, a child with a complexity limit of 2 will only be able to apply inversion in affirmative WH-questions, and so on.

Unfortunately, without further adjustments, this kind of explanation predicts that the following types of structure should also be produced by children at the appropriate stages of acquisition:

(60)a. *Did you see what ?
b. *You didn't see what ?
c. * Didn't you see what ?

But as we have seen, in-situ WH-questions are never produced by children at the relevant stages of development; moreover, this is not due to any difficulty of interpretation, since, as we saw in 5.33, K. had no trouble imitating or interpreting such questions.

Of course, this is a familiar problem - and not an irremediable one - for acquisition explanations based on some index of cumulative complexity. For example, the Basic Operations Hypothesis of Mayer et al. (1978), discussed in Chapter 3, faces exactly the same problem of "error overgeneration". There are various ways around the problem: the most popular is to stipulate that only certain syntactic

process can be "undone" in case the complexity limit is exceeded (Goodluck and Solan 1979, for example, exempt rules which contain essential variables - i.e., WH-movement). However, the question then arises as to whether such a construction-specific approach to complexity limitations has any explanatory potential, since there appears to be no built-in constraint on which rules should fail to apply when the limit is exceeded. For example, in order to make this approach work for SAI development, it would have to be stipulated that WH-movement, SAI and negation all "counted" for cumulative complexity, but only SAI could be "undone" in case the limit was exceeded in a given sentence.

This in turn raises the larger question of the relation of a cumulative complexity index to syntactic theory. When Bellugi first proposed her complexity-based explanation, matters were reasonably simple, since the only rule-types available were context-free phrase-structure rules and transformations; but since then, the theory of grammar has changed rather radically. For example, the three transformations of negation, SAI, and WH-movement on which Bellugi based her original account of SAI development are treated very differently in contemporary syntactic theory: WH-movement is a successive cyclic A'-movement rule creating an operator-variable structure at LF; SAI is a bounded head-movement rule, motivated according to our account by conditions on predication; and negation no longer involves movement at all, since its scope is determined by structural conditions applying at S-structure. In the face of

this heterogeneity, it appears extremely unlikely that any simple additive measure of cumulative complexity will be able to tell us much about acquisition sequences.

In the face of the problematic relationship between complexity measures and syntactic theory, one move might be to divorce the two altogether. However, this creates far more difficulties than it resolves. In particular, it is clear that complexity limitations of the type discussed by Bellugi cannot be reduced to a more general measure of "cognitive complexity", since they are intimately tied to particular syntactic constructions. Thus WH-questions with and without inversion are semantically identical for the non-inverting child - whatever prevents inversion cannot be defined in terms of interpretive difficulty.

Thus we are left with an obviously syntactic notion of cumulative complexity, but one which does not appear to translate into any particular syntactic framework. In the face of this, we can make one of two moves; abandon the complexity hypothesis, or abandon the syntactic framework. The decision is an empirical one, and it seems clear to me that empirical considerations favour the former option.

Once again, take Bellugi's SAI explanation as an example. It turns out that even when restricted to the narrow data-base provided by the Harvard study, her explanation only partially handles the particular acquisition sequence which characterizes the SAI development of the Harvard children; while negation appeared to affect the probability of inversion in WH-

questions, she reports that it had no inhibitory effect on yes-no questions (see Erreich 1984, though, for a different view). Empirical problems multiply when the data-base is broadened; the complexity hypothesis clearly cannot be extended to deal with SAI developmental patterns such as those exemplified by S. or K., as detailed in 5.32 and 5.33 above, nor the non-inversion patterns reported by Ingram and Tyack (1979) or Erreich (1984).

We conclude that since the complexity hypothesis is both theoretically suspect and empirically unsuccessful, it should be abandoned. Accordingly, we turn to a quite different type of explanation for SAI development, this time based on the properties of intermediate grammars, and exemplified in the rather different accounts provided by Klein (1982) and Pinker (1984). It is Klein's account which we will turn to first.

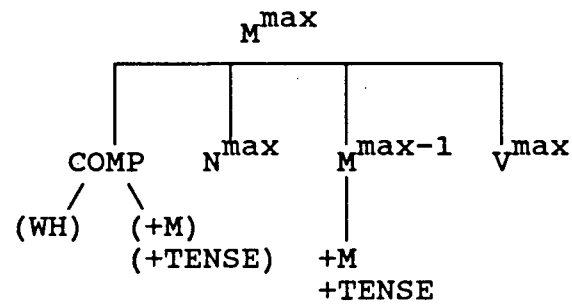
5.62 Klein's account of SAI development

Just like Bellugi's account, Klein's is interesting in that it represents very clearly a particular style of explanation in language acquisition. This time, however, it is a purely grammar-based type of approach; in fact, Klein's work is a pioneering attempt to apply the linear-parametric model of syntax to data from language acquisition, and as such is particularly interesting from the perspective of our investigation into the relation between syntactic theory and real-time language acquisition.

Let us recall once more the principle characteristics of the linear-parametric model. The first language learner is equipped by UG with a set of principles, some of which are subject to limited parametric variation; values of a parameter are arranged in order of markedness. The Subset Principle of Berwick (1982) ensures that at least some of these values are intrinsically ordered by "delearnability" considerations (see Chapter 1 for details), but it is unclear whether all values can be ordered in this way. During language acquisition, the child will first attempt to apply the least marked value of a parameter to the language s/he is learning; if positive counter-evidence is encountered in the PLD, s/he will move on to the next least marked value, and so on until no more counter-evidence is encountered.

Now, as pointed out frequently in this thesis, the linear-parametric model predicts that if a language sets a parameter for a markedness value of greater than 0, we will expect to find "stages" in language acquisition corresponding to each parameter-setting, running from least to most marked. And indeed, this is exactly what Klein claims for the acquisition of SAI. Accordingly, let us examine the "intermediate grammars" which she proposes, beginning with G1, corresponding to Period B in Bellugi's account of SAI development (from which Klein takes most of her data):

(61)



The principle differences between G1 and the adult grammar are the following:

(a) There is no S' (CP) constituent; COMP is a branching node generated under S, which is headed by M (=Modal).

(b) AUX elements are base-generated in both pre-sentential and pre-verbal positions.

(c) WH-phrases are base-generated in COMP.

These differences allow Klein to account for the following (purported) characteristics of Period B:

(i) Apparent cases of SAI.
(ii) Auxiliary overmarking errors.
(iii) The absence of WH-copying errors and echo questions.

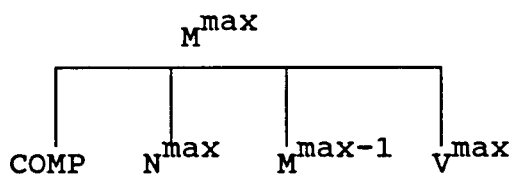
In fact, it can be argued that none of these properties should be captured through an intermediate grammar. Apparent cases of SAI at this stage, as we have seen, are indeed only

apparent; they involve formulaic routines, which should not be represented grammatically. Auxiliary overmarking errors are not characteristic of Period B; they occur at all stages of acquisition, and surface as adult speech errors. Finally, as we have pointed out, the absence of WH-copying errors and echo questions is a non-issue in language acquisition, as long as we bear in mind that the important aspect of WH-constructions is the filler-gap dependency, not the source of the filler.

G1 is not only empirically inadequate, but also rather peculiar in terms of the theory of grammar. Since it is the language learner's first hypothesis, it should have a markedness value of 0; however, it is certainly <u>not</u> the case that grammars which generate two filled tensed auxiliary positions in the same clause are unmarked (in fact, I don't know of any at all).

Let us then turn to G2, which is meant to represent Bellugi's Period C:

(62)



The following changes have occurred:

(a) AUX is no longer generated in COMP, and COMP is no longer branching.

(b) The child has postulated a COMP-substitution rule, formulated as in (63):

(63) [COMP, [+WH]] - X - AUX ---> 3 - 2 - 0
1 2 3

G2 is designed to account for the following Period C characteristics:

(i) Inversion in yes-no questions.
(ii) Non-inversion in WH questions.

The main problems with G2 (aside from the prediction that all children will go through a Bellugi-type Period C, which is false, as we have seen) concern the type of evidence needed to effect the transition from G1 to G2. The child has to learn first that AUX can no longer be generated in COMP, and second that COMP is non-branching, but there does not appear to be positive evidence for either transition in the PLD; indeed, it is difficult to see how the child could ever retreat from the overgeneral grammar G1 to the restricted grammar G2 without employing negative evidence.

Neither of the solutions Klein offers to these two transition problems is convincing. She suggests that learning the relevant features of AUX(M) might trigger retreat from the base-generated AUX-in-COMP hypothesis, but unfortunately does not specify which features or how they might effect the change. Her solution for the transition from a branching to a non-

branching COMP is equally vague and ad-hoc: she speculates that the failure of sentence-initial NEG in COMP to c-command its scope in S might be the relevant trigger, but does not specify how the child realizes that S-initial NEG fails to c-command its scope in the first place, a type of knowledge which surely cannot be provided by positive evidence.

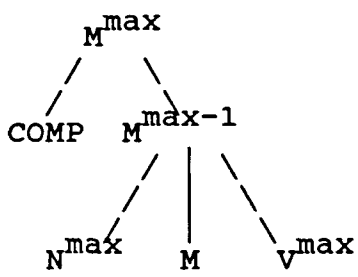
A final incorrect prediction made by G2 is that auxiliary overmarking errors will disappear during Period C. But of course, they occur; here are two from K.:

(64)a. Why is there's big tears ?

b. What is the woman is doing ?

Let us then turn to G3:

(65)



G3 differs from G2 only in that COMP is repositioned outside S; evidence for this is provided by extraction from embedded clauses, which must involve successive cyclic movement and therefore a COMP "escape hatch". Of course, it must be stipulated that such evidence only becomes available at this point, rather than earlier; presumably, Klein would appeal to

performance factors limiting the accessibility of degree-1 PLD in order to achieve this.

Finally, the child must reformulate the substitution rule in (63) as an adjunction, in order to pass out of the non-inversion stage. Of course, SAI in WH-questions provides the right evidence, but once again, one wonders why it only comes to be used at this late stage in syntactic development.

While Klein's study suffers from a number of empirical and theoretical problems, it seems to me instructive to sort them out into those which are specific to her analysis and those which apply in principle to <u>any</u> linear-parametric treatment of SAI development. The latter class is intrinsically more interesting, because it reveals the inherent limitations of the linear parametric model when faced with data from real-time language acquisition.

Probably the most salient problem of the latter type is that posed by the existence of "stages" in acquisition - such as the non-inversion stage - which characterize only a small proportion of learners. There are ways round this - it could be claimed that most children pass through the non-inversion period "silently", for example - but such "solutions" are unmistakably ad-hoc.

Another problem concerns variation between children. We have seen that by no means all children who make non-inversion errors follow the path traced by Bellugi - in fact, Bellugi's acquisition sequence seems to characterize a minority of non-inverters, who in turn constitute a minority of childrem

learning SAI. The linear-parametric model has no way in principle of dealing with children who appear to pass through different intermediate stages in the acquisition of SAI, without hypothesizing different parameter-settings for different children, a move which would undermine the very foundations of the parametric model.

A third problem concerns the mis-match between errors in acquisition and the theory of syntactic markedness, discussed in Chapter 1 with reference to the parametric account of the "missing subject" stage and exemplified particularly clearly by KLein's G1, which is not only marked but probably impossible as an adult grammar. In fact, the linear parametric model predicts that children's syntactic errors will always go in the direction of a less marked grammar, a prediction which appears to be falsified by data from real-time language acquisition. To give another example, we saw in Chapter 4 that a language will only have inversion in yes-no questions if it also has inversion in WH-questions, but not vice-versa (as expressed by Word Order Universal no.11 in Greenberg 1963); however, the non-inverting stage - at least, the type described by Bellugi - appears to go in the opposite direction, with inversion in yes-no questions being apparently easier to acquire than inversion in WH-questions. We will return to this problem in 5.7, where we will argue that "transitional" grammars in acquisition are invariably marked with respect to the adult grammar.

Finally, there is the problem of triggering evidence. Every linear parametric account of acquisition sequences is forced to

make the assumption that relevant evidence for the transition from one intermediate grammar to another is available, but ignored by the child up until the transitional point; yet no account has any suggestion as to why this should be the case. Without the rudiments of a theory of "intake" - i.e., what type of evidence the child uses when, and why - the availability of PLD simply becomes a deus ex machina which robs the linear parametric model of much of its explanatory potential.

Bearing these problems in mind, let us then turn to our second grammar-based account of SAI acquisition, that of Pinker (1984).

5.63 Pinkers's account of SAI acquisition

The central component of Pinker's explanation for the developmental sequences characteristic of SAI learning is the idea, first proposed by Grimshaw (1977), that inversion and WH-movement are for the most part independent. Of course, this theoretical position is the very opposite of that which we have adopted in our treatment of SAI-type inversion rules; this makes Pinker's developmental account particularly interesting from the point of view of the instantiation of specific syntactic hypotheses in models of language acquisition.

In the LFG theory adopted by Pinker, each WH-word is supplied in the lexicon with a constraining equation for "sentence-modality" (it should be emphasized that the latter

notion is strictly syntactic in Pinker's account). In the case of a WH-phrase which induces inversion, the equation will be represented as in the lexical entry (66a.); in the case of a WH-word which does not, the equation will be represented as in (66b.):

(66)a. what PRED = "for which X"
wh = +
SENT-MOD not $=_c$ noninv

b. how come = "why"
wh = +
SENT-MOD $=_c$ noninv

(Note that the feature "noninv" rather than "inv" is employed in order to avoid a technical problem concerning constraints on feature percolation in LFG, which need not concern us here).

In order to avoid inversion in non-root clauses, Pinker employs an additional constraining equation, ROOT $=_c$ ±, whose value constrains the category symbol on the right-hand side of the rule expanding S; if the category symbol is (V) (i.e., the clause is verb-initial, then the value must be [+], if not, it must be [-].

This theoretical apparatus allows Pinker to produce a "lexically-based" account of SAI development; in particular, since learning SAI involves learning the constraining equation on each WH-word, the "non-inversion stage" can be accounted for simply by claiming that the non-inverting child has not learnt the correct equation for a given WH-word. Since in the syntactic model Pinker adopts, WH-movement and inversion are unrelated in the unmarked case, this means that the child who

has not yet received enough information to hypothesize the correct equation will treat inversion and "WH-movement" as independent processes; since there is an optional phrase-structure rule for both sentence-initial WH-phrases and fronted auxiliaries (recall that there are no movement rules in LFG), this in turn predicts that inversion should be optional for the non-inverting child.

While ingenious, this account is flawed in at least three ways. First of all, it makes incorrect predictions about cross-linguistic markedness; second, it makes the wrong predictions about acquisition sequences; and third, it encounters serious learnability problems. Let us turn in more detail to each of these issues.

First of all, let us deal with the issue of markedness. As noted above, in order to make his acquisition model work, Pinker is forced to claim that in the unmarked case "WH-movement" and inversion are independent; moreover, since children make non-inversion errors, the marked case, where the two processes are interdependent, must be relatively inaccessible to the first language learner. Pinker does in fact attempt to support this claim by citing Slobin's (1985) cross-linguistic generalization to the effect that children conform to the tacit injunction that "an operator that affects a whole phrase or clause should not be placed within that phrase or clause, nor should it require the changing of elements within that phrase or clause." However, it is difficult to see exactly where such a generalization fits into the acquisition theory:

is it, for example, an effect of the acquisition process itself, and thus specific to children, or is it meant to characterize adult grammars as well ?

In fact, it is possible to show that the first of these alternatives is theoretically incoherent, and the second empirically wrong. Suppose we were to claim that Slobin's generalization was restricted to child language. In that case, we could adopt either of two positions: we could claim that it was not part of the grammar at all, but in effect caused by some kind of interference from other cognitive systems, or we could claim that it was part of some grammatical markedness heirarchy specific to language acquisition. The former position is in fact equivalent to that taken by Bellugi, which we have already discussed in 5.61, and suffers from the same kind of difficulty - namely, the problem of how to define "cognitive complexity" in a way which is language-independent, yet appears to be tied to specific syntactic constructions. The other possibility, on the other hand, sets up a "developmental acessibility heirarchy" of syntactic hypotheses which bears no relation to their distribution in adult grammars; such a move effectively undermines any approach to language acquisition - including Pinker's own theoretical framework - which attempts to relate syntactic markedness to markedness in acquisition.

Suppose then, we were to claim that Slobin's generalization held for adult as well as developing grammars. In that case, we would expect cross-linguistic investigation to show that grammars where WH-movement and SAI are interdependent are

marked relative to those where the two processes act independently. But as we have pointed out a number of times, precisely the opposite situation in fact holds: if a grammar has yes-no question inversion, then it will have WH-induced inversion, but not vice-versa. Thus we see that the accessibility heirarchy which Pinker is forced to set up in order to explain developmental markedness cannot be generalized to deal with cross-linguistic markedness. Of course, this is exactly the same "non-correspondance" problem which we discussed in relation to Klein's linear-parametric account of SAI acquisition in 5.62, and which we will return to in 5.7.

Let then us turn to the empirical flaws in Pinker's account. Here, the problems are precisely the opposite from those encountered by the linear-parametric model. Whereas the latter predicts a set of rigidly ordered universal stages, Pinker's explanation predicts no order at all; essentially, anything goes at the non-inversion stage. In fact, Pinker cites the work of Ingram and Tyack (1979), Kuczaj and Brannich (1979), and Erreich (1984) in support of this claim.

Now, whereas none of these studies show the invariant stages predicted by the linear-parametric model, they do show regularities which cannot be accounted for simply by assuming that inversion is optional. For example, Pinker claims that the overgeneralization of how come in the Kuczaj and Brannick study lends support to the optionality of inversion in children's language at this stage, but as we have seen, overgeneralization with how come only occurs after the non-inversion stage.

Likewise, Pinker's account predicts that we should get any combination of inversion possibilities in the language of non-invertesr, but we have shown that at least one (involving inversion in WH- but not yes-no questions) does not occur. Finally, Pinker has no account of why only a small proportion of children should end up as non-inverters - a problem shared by all the other SAI explanations we have reviewed.

However, the most serious problem of all is not empirical in nature, but concerns the "delearnability" of the "optional inversion" stage. Given the standard assumption that negative evidence is unavailable to the first language learner, it is easy to see that a child who has postulated an optional rule for inversion in WH-questions will never be able to retreat from that hypothesis, since no positive evidence will ever tell him or her that a non-inverted question is ungrammatical. In other words, Pinker's account is caught in a classic learnability trap of the type first described by Baker (1979).

We conclude that Pinker's explanation for SAI development fails to stand up to close scrutiny on three counts, involving markedness, acquisition sequences, and delearnability, and must therefore be rejected.

5.64 Input-based accounts of SAI acquisition

We now turn from grammar-based accounts of SAI development to those based on features of the child's input. The latter type of explanation has been less popular and less well-

articulated than the former, so our discussion will be briefer; nevertheless, we will argue that some aspects of the child's input are crucially important in accounting for SAI developmental sequences, and will play an important part in our own account in 5.7.

One input-based explanation for non-inversion errors, put forward by Brown and Hanlon (1970) and Maratsos (1978) is that embedded clauses, where inversion normally never takes place, could be providing misleading information for the non-inverting child. However, such an explanation leads to some implausible assumptions about the nature of the PLD available to the first language learner; in particular, it implies that non-inverting children must be using degree-1 evidence in formulating their grammars for SAI. Of course, we cannot rule this possibility out a priori; however, as we shall see, it turns out that empirical considerations independently argue against the child's use of such evidence.

There are two ways in which embedded clauses might count as evidence for non-inversion. First, the child might use both degree-0 and degree-1 evidence: i.e, any clause which had undergone WH-movement would count as relevant PLD. Alternatively, the child might <u>only</u> use degree-1 evidence, selectively attending to subordinate clauses. Let us examine these alternatives one by one.

Suppose any kind of WH-question, embedded or matrix, counted as PLD. In that case, we would expect not only non-inversion errors in matrix clauses, but at the same time

inversion errrors in subordinate clauses. But as we have seen, the latter type of error, if it occurs at all, only occurs after the non-inversion stage; we conclude that a learner employing both degree-0 and degree-1 evidence would not show the error-pattern typical of the non-inversion stage.

Alternatively, suppose the child was programmed to attend only to embedded clauses (again, we cannot dismiss this possibility a priori, implausible as it may seem). This would solve the problem alluded to above, but it would be incompatible with the observation, made in 5.33 above and by Phinney (1981), that children at the non-inversion stage appear to have difficulty with learning some of the distinctive properties of embedded clauses, in particular, the complementizer that, which is neither employed in spontaneous production, nor imitated correctly. In order to account for this, a non-inversion account based on selective attention to subordinate clauses would have to propose that the non-inverting child was capable of picking out subordinates (as opposed to coordinates, for example), but incapable of employing the single most salient clue to subordinate status - the lexical complementizer. While not beyond the bounds of posssibility, this seems to stretch plausibility to breaking point. When coupled with the fact that many non-inverting children do, of course, produce inverted questions - indicating that matrix questions must be playing some role in the PLD of non-inverting children - this evidence indicates that the "degree-1 evidence only" hypothesis cannot be maintained.

Thus we see that neither interpretation of the claim that embedded WH-questions are responsible for misleading the non-inverting child stands up to careful scrutiny.

There is, however, another way in which the PLD might mislead the child learning SAI, which as far as I know has not been previously discussed. This is the possibility that unanalyzed (WH + contracted auxiliary) sequences might provide evidence that inversion is not required in WH-questions. This is by no means an implausible suggestion, since it is well-known that the phonological relationship between contracted and non-contracted forms of the auxiliary in English is highly opaque; indeed, Zwicky and Pullum (1983) and Kaisse (1985) argue that the two sets of forms are not phonologicaly related at all, on the basis of the non-existence of independently motivated phonological rules which could derive the contracted from the non-contracted form. Thus, there is no rule deleting [w] in English which could derive [l] from will and [d] from would, for example. Moreover, even if such a rule were to be postulated solely on the basis of "auxiliary reduction", it would have to be stipulated that it could not apply to the auxiliary forms was and were, which have no reduced variants, as shown by the examples in (67):

(67)a. He was going. --/--> He's going.

b. They were going. --/--> They're going.

Other reduced and non-reduced auxiliaries have partially

distinct syntactic distribution; the contracted form 's occurs in environments where its full form is ungrammatical, as shown in (68) (examples from Kaisse 1985):

(68)a. *Where is the lions ? --/--> Where's the lions ?

b. *There is/has a new book been written. --/--> There's a new book been written

Finally, Akmajian, Demers, and Harnish (1984) motivate an auxiliary reduction process ("Tag-controlled deletion") which deletes an entire (subject NP + contracted auxiliary) sequence under identity with a tag pronoun:

(69)a. (You've) been hitting the bottle again, haven't you ?

b. (You)* have been hitting the bottle again, haven't you ?

c. (You'll) steal my money, will you ?

d. (You)* will steal my money will you ?

e. (It) could be interesting couldn't it ?

f. *(It'd) be interesting, couldn't it ?

The most interesting (and peculiar) aspect of the sentences in (69) is that in order for Tag-controlled deletion to apply, if an auxiliary is in the main clause, it must contract to the subject NP just in case it can so contract. Now, if auxiliary reduction was phonological, Tag-controlled deletion, which is sensitive to its output, would also have to be phonological - but this is stretching the notion of a phonological process beyond the bounds of credibility, since Tag-controlled deletion

is an obviously syntactic process involving the deletion of a subject under identity with a remote controller.

Given these arguments, it appears that contracted auxiliaries must be analyzed as non-phonological clitics (this is the conclusion reached by both Zwicky anmd Pullum and Kaisse). It would therefore be hardly surprising if children learning SAI were to be misled into believing such clitics bore no syntactic relation to the full forms of auxiliary verbs (as attested in yes-no questions), but instead were agreement markers affixed to NPs. Moreover we have seen evidence that non-inverting children do fail to properly analyze sequences involving both (subject NP + contracted auxiliary) and (WH-phrase + contracted auxiliary); K., for example, produced sentences such as those in (70):

(70)a. How much spikes do they're have ? (39 mos., MLU 3.72)
b. What's that noise is ? (37 mos., MLU 3.38)
c. Yes, they're can. (37 mos., MLU 3.38)
d. What's these are ? (35 mos., MLU 3.56)

Thus, we conclude that the PLD in English is indeed capable of "misleading" the non-inverting child; in the next section, we will make use of this type of "input confusion" in our own account of the developmental sequences characteristic of SAI acquisition.

5.7 An Integrated Account of the Acquisition of SAI

5.70 Introduction

In this section, I will attempt to put together the various pieces of the puzzle I have been assembling throughout this chapter. I will propose an explanation for SAI acquisition which takes into account the following characteristics of the developmental sequences we have been examining:

(i) In contrast to the acquisition of INFL2 in Germanic languages, the development of SAI is slow and uneven.

(ii) Nevertheless, only 25-30% of children learning English pass through a "non-inversion" stage.

(iii) Of those children who pass through the non-inversion stage, some invert in yes-no questions but not in WH-questions, while others invert in neither.

(iv) The learning of SAI in WH-questions is subject to both type- and token- variability.

(v) Limited overgeneralization of SAI with <u>how come</u> sometimes takes place after the non-inversion stage.

(vi) Limited overgeneralization of SAI to embedded clauses also takes place during a period after the non-inversion stage;

children learning INFL2 apparently never make this type of error.

(vii) Negation appears to have an inhibitory effect on SAI.

THe essential ingredients of my explanation will be the following:

(a) The AGR-based account of SAI-type inversion rules developed in Chapter 4.

(b) The open/closed class distinction, as encoded by Emonds' (1985) Designation Convention.

(c) A PDP-type learning mechanism of the type described in Chapter 1 and employed elsewhere in this thesis.

My strategy will be the following: in 5.71, I will compare the developmental patterns exemplified by children learning SAI and INFL2, accounting for the similarities by (a) and the differences by (b). In 5.72, I will turn to the "non-inversion stage", arguing that it does not constitute a stage at all, but a particular kind of "acquisitional misprojection". In 5.73, I will deal with the question of the relation between yes-no questions and WH-questions in the non-inversion stage, which in turn will lead me to propose a rather different notion of "transitional grammar" than those currently in use. In 5.74, I

will examine the question of lexical effects and variability in SAI acquisition, arguing that (b) and (c) above can be used to construct an effective account for the observed patterns of development. Finally, in 5.75, I will discuss some residual problems in the explanation of SAI development.

Let us then proceed with the account.

5.71 SAI, INFL2, and AGR

In 5.5, we saw that the developmental patterns characteristic of the acquisition of INFL2 languages differ in many ways from the SAI acquisition sequences we have been investigating in the rest of this chapter. However, before turning to these differences, I would first like to explore the implications of one striking <u>similarity</u> between the two processes.

This is the fact that children learning SAI and INFL2 both begin by producing clauses which have been "de-inverted". While in the case of SAI, this could be attributed to the greater input frequency of sentences with "declarative" word order (but see Newport, Gleitman, and Gleitman 1977, who argue on the contrary that <u>inverted</u> sentences are more common in parental speech to children) this certainly cannot be the case in INFL2 languages, where inversion is obligatory in <u>all</u> main clauses. As we argued in 5.5, this seems to indicate that the child learning INFL2 must already know how the rule works in order <u>not</u> to apply it, an apparently paradoxical conclusion.

In fact, Newport, Gleitman and Gleitman (1977) provide some

interesting evidence that appears to force us to come to exactly the same conclusion about the learning of SAI. They show that the learning of auxiliary verbs in medial position appears to be contingent on input containing yes-no questions, where auxiliaries are in a phonologically more salient fronted position. Now, obviously, in order to use this type of input, the child must make the connection between auxiliaries in fronted and medial positions; in other words, the child must be able to figure out the rule of SAI. However, at the period in question, children do not produce inverted questions, but instead rely entirely on declarative word order with question intonation. In other words, we are forced to the same conclusion that we came to regarding the learning of INFL2: the child must already know the rule of SAI, in order to be able not to apply it.

Further support for this conclusion is provided by Kuczaj and Maratsos (1983), in their discussion of the learning of yes-no questions. They argue that during a developmental period roughly corresponding to MLU range 3.0 - 4.5 (i.e., exactly the same period as the non-inversion "stage"), children who correctly use auxiliaries in both declaratives and yes-no questions often fail to relate the two positions. Two types of evidence support this contention: first of all, such children may employ a given auxiliary in medial position without ever generalizing it to the fronted yes-no position (as in the case of Kuczaj's son Abe, who used will, can, would and could in declaratives but only will and can in yes-no questions);

second, as is well-known, children who use lexically restricted auxiliaries such as better and catenatives such as gonna and hafta never overgeneralize these elements from medial to fronted positions - errors such as those in (71) are not attested in the child language literature:

(71)a. *Better I go home now ?
b. *Hafta I go to bed ?
c. *Wanna you drink some tea ?

Now, as argued above, such children do know that in general auxiliaries occur in either fronted or medial position - i.e., they have a general knowledge of the existence of SAI. Yet once again, they apply this knowledge only partially and in a lexically restricted fashion.

What this suggests - and what I will propose in more detail in 5.8 - is that syntactic learning takes place in two distinct modes. The first mode involves very general, rapid acquisition of the fundamental syntactic properties of the language to be learnt, and takes place at a comparatively early stage in syntactic development; the second involves the slow, uneven, variable and lexically restricted kind of learning which we have seen to be characteristic of the later development of SAI. One way of looking at this kind of developmental schema is to imagine that children at the early stages of acquisition are sketching out the broad outlines of the language they are learning, and only later do they turn to filling in the

details.

Now, the question arises as to what exactly triggers the switch from "general" to "specific" learning. I would like to suggest that it is the learning of closed-class categories which acts as the trigger; the role of such categories is precisely to mediate between general syntactic rules and specific lexical items. Again, this proposal will be discussed in more detail in 5.8, where I will provide a more articulated "two-tiered" model of the acquisition of syntax.

Bearing all this in mind, let us then return to the acquisition of INFL2 and SAI. We have seen that at the beginning of the developmental sequence, children learning the two processes act much alike: they both make the generalization relating fronted and medial finite verb positions, but neither employs inversion. In terms of the two types of learning we have been discussing, type-1 (general) learning is then similar for both INFL2 and SAI.

However, as we have seen, after the early stages of acquisition, children learning INFL2 and SAI begin to diverge significantly. INFL2 learners make a very rapid transition from the non-inverting to the inverting stage, contingent on the learning of agreement morphology; SAI learners go through the prolonged, uneven and variable developmental period which we have been investigating throughout this chapter. Thus type-2 (specific) learning differs markedly in INFL2 and SAI acquisition.

We are now in a position to account for both the

similarities and the differences between the acquisition of SAI and INFL2: while the former derive from the syntactic principles which underly both processes (as argued in detail in Chapter 4) the differences spring from the different types of learning necessary to acquire the two systems. Let us deal with each in turn.

Recall first of all the major aspects of our account of SAI-type inversion, as detailed in Chapter 4. There, it was argued that inversion was forced by A'-type predication, through the P-licensing condition in (72) and the AGR-linking condition in (73):

(72) P-licensing condition III:

Every P(redication)-subject must be AGR-linked.

(73) AGR-linking:

A P-subject α is linked to AGR_i iff α minimally governs AGR_i.

(See Chapter 4 for details.)

Now, we will assume that (72) and (73) are part of the child's innate syntactic knowledge - that is, they are part of UG. In that case, "learning" an SAI-type inversion rule will consist of solving two problems: firstly, the learner must recognize that A'-type predication is instantiated in the language to be learnt; and secondly, the child will have to identify the domain of its application - i.e, which elements count as P-subjects, which syntactic elements carry AGR, and which configurations meet the structural conditions on AGR-linking.

I would like to suggest that these two problems - which I will call the recognition and instantiation problems - correspond to type-1 (general) learning and type-2 (specific) learning, respectively. Thus at an early stage of acquisition, children learning both SAI and INFL2 recognize that they are dealing with a case of A'-type predication and know (via UG) which principles and parameters apply - that is, they have solved the recognition problem; however, they do not know exactly how those principles and parameters are instantiated; in other words, they have not yet come to grips with the instantiation problem.

Next, let us return to the question of what the instantiation problem involves in INFL2 and SAI languages, respectively. As pointed out above, in both types of language, the child will have to learn which elements bear AGR, which elements count as P-subjects, and which configurations meet the structural conditions on AGR-linking. We will assume that once the phrase-structure of the two types of language has been deduced (see Chapter 2 for ideas on how) the latter problem willl be resolved; this leaves the questions of which elements bear AGR and which can count as A'-type P-subjects.

Here we get to the crux of the difference between INFL2 and SAI learning; solving the instantiation problem for the child acquiring INFL2 need only involve a small amount of closed-class learning (the acquisition of the relatively transparent inflectional morphology of subject-verb agreement); on the other hand, the child acquiring SAI must use closed-class

learning not only to determine the morphology of the agreement system in English (which is itself relatively opaque) but also to determine which elements may bear AGR (and thus undergo SAI) and which elements act as A'-type P-subjects (and thus induce SAI).

We can now account for the relatively rapid transition from the non-inverting (or rather, pre-inverting) stage in INFL2 acquisition to the inverting stage, as well as for the correlation between INFL2 acquisition and the acquisition of subject-verb agreement morphology. In order to instantiate the P-licensing and AGR-linking conditions, the child learning INFL2 must acquire the specific closed-class elements (i.e., verbal inflections) which bear AGR; however, once this task is accomplished, no more closed-class learning is necessary, since any XP whatsoever can act as an A'-type P-subject, and any finite verb can bear agreement morphology. Thus, once agreement morphology is learnt, the child can immediately generalize INFL2 to any member of the open classes to which it appplies.

On the other hand, the child learning SAI is faced with a lexically restricted process which requires closed-class learning of at least three kinds:

(i) Learning inflectional morphology.

(ii) Learning A'-type P-subjects.

(iii) Learning AGR-bearers.

Thus the difference between the developmental patterns characteristic of INFL2 and SAI acquisition can be traced to the types of learning necessary to acquire the two systems: INFL2 acquisition involves only a small amount of closed-class learning, while SAI requires a large dose.

We now have an account of the first of the seven characteristics of SAI development which we presented in 5.70 above. In the next section, we will turn to the second - the question of how the types of learning in (i - iii) above might lead the child to enter the non-inversion stage in SAI acquisition.

5.72 SAI Development and the Non-Inversion Stage

In this sub-section, we will focus on the non-inversion "stage" in SAI development. The central idea behind our account will be that non-inversion is not a stage at all, but, in effect, a "garden path"; we will argue that the child learning SAI may be misled by certain aspects of the PLD into adopting an analysis in which the P-licensing condition in (72) is satisfied without resort to SAI.

The key to the non-inversion "stage", I would like to claim, is the misanalysis of (WH-phrase + contracted auxiliary) sequences by non-inverting children. As argued in 5.64 above, this is a plausible type of mistake: the relation between reduced and non-reduced auxiliaries is opaque, and in WH-questions, auxiliaries occupy "Wackernagel's position", the

prosodically weak position which follows the first streesed constituent in the sentence and is associated cross-linguistically with sentential cliticization (see Kaisse 1985). Moreover, it is known that children at the non-inverting "stage" fail to segment sequences containing reduced auxiliaries; and in our investigation of SAI development we have seen that at certain periods WH-questions with contracted auxiliaries are the most frequent type of question produced by children learning SAI: for example, in K.'s Period II, they outnumber all other question-types put together.

What I would like to suggest is that those children who misanalyze (WH + contracted auxiliary) sequences reanalyze them as (WH + AGR-clitic) sequences; in other words, they understand that cliticized auxiliaries satisfy the AGR-linking condition, but fail to realize that the cliticized forms are in fact auxiliaries. Now, this predicts that SAI need not take place in order to link the A'-type P-subject (i.e., the WH-phrase) to AGR by (73), and thereby satisfy the P-licensing condition in (72), since the WH-phrase already has an AGR-clitic attached to it - hence, non-inversion.

One interesting prediction made by this analysis is that while we should expect errors involving WH-questions containing contracted auxiliaries in fronted position as well as auxiliaries in medial position, we should <u>not</u> expect to find WH-questions containing contracted as well as uncontracted auxiliaries in fronted position; if, on the other hand, children who misanalyze (WH + contracted auxiliary) sequences

simply treat them as variants of WH-phrases, the latter type of error should be as frequent as WH-questions with inverted auxiliaries. In fact, I no of no instances of the latter type of error; K., for example, produced WH-questions such as those in (74a.-c.) at the same point in his development (Period II, when he was 35 months old and had an MLU of 3.56) but never produced questions such as that in (74d.):

(74)a. What is this ?

b. What's this is ?

c. What's that ?

d. *What's is that ?

This appears to me to provide rather striking supporting evidence for the analysis we have adopted.

Another virtue of this analysis is that it enables us to account for the fact that only 25-30% of children learning SAI make non-inversion errors; this is because only that proportion of children make the original segmentation error which leads them to reanalyze (WH + contracted auxiliary) sequences as (WH + AGR-clitic) sequences. In other words, by tying non-inversion to a rather low-level failure of morphophonological segmentation, we have avoided the problems attendant on postulating universal "stages" in syntactic acquisition which turn out to characterize only a minority of language learners. Moreover, since children acquiring SAI are engaged in closed-class learning, we predict that non-inversion errors should

exhibit the typical cluster of properties associated with such learning, notably "lexical" effects and type- and token-variability; and indeed, we have seen ample evidence for such effects in our investigation of SAI development.

We have thus accounted for the second of the seven characteristic properties of SAI development which we presented in 5.70: since the non-inversion "stage" is not a stage but a misprojection, we do not expect every child to make non-inversion errors.

One very important outstanding issue remains, however; while we have accounted for non-inversion in WH-questions, we have not yet touched upon the relation between non-inversion in WH-questions and in yes-no questions. This will be the subject of the next sub-section.

5.73 The Interaction of Yes-No and WH-Questions in the Acquisition of SAI.

It might very well be asked at this point why the phonologically salient, uncontracted, inverted auxiliaries in yes-no questions do not immediately provide the non-inverting child with evidence for SAI.

The obvious answer to this question is that yes-no questions and WH-questions are treated independently by the language-learner, and therefore evidence for inversion in one type cannot serve as evidence for inversion in the other.

However, two considerations prevent us from adopting this

position. Firstly, as we have pointed out throughout this chapter, there is a cross-linguistic markedness relation between inversion in yes-no and in WH-questions: if a language has the former, it will have the latter, but not vice-versa. We will for the purposes of the current discussion treat this markedness relation as a primitive property of UG; though ideally it should be made to follow from more fundamental principles (see Chapter 4 for some suggestions) nothing in the arguments given below will hinge on the issue.

Second, if children were to treat SAI in the two types of question separately, we would expect children learning SAI to show all four possible combinations of inversion and non-inversion; however, we have argued in 5.3 that one pattern is missing: that involving inversion in WH-questions but not in yes-no questions.

It seems, then, that there is an interaction between the two types of question in the development of SAI. In this sub-section, we will explore this interaction in more depth, arguing that the type of "transitional" grammar exemplified by non-inverting children is created by the need to resolve the conflicting demands of UG and the PLD; in order to do so, no-inverters are forced either to ignore the PLD or temporarily adopt an "impossible" grammar. Since this situation is inherently unstable, the non-inverting "stage" contains the seeds of its own decay; we do not need to appeal to the sudden appearance of triggering evidence or other maturational considerations in order to explain how the non-inverter

eventually acquires the adult grammar of SAI.

We have identified two types of non-inverting child. The first, exemplified by the Harvard children, inverts in yes-no but not in WH-questions; the second, exemplified by K., inverts in neither. Let us assume that the reason for non-inversion in WH-questions (which characterizes both sets of children) is that argued for in 5.72 above: namely, the misanalysis of (WH + contracted auxiliary) sequences as (WH + AGR-clitic) sequences.

Now, as argued above, inverted auxiliaries in yes-no questions are not subject to the same kind of misanalysis; indeed, of all the possible auxiliary positions, the fronted yes-no position is the phonologically most salient (which explains its importance in learning auxiliaries in the first place, as observed in 5.71 above).

The child who has misanalyzed WH-questions is then faced with a prima facie contradiction. UG (in the form of the markedness statement relating inversion in yes-no and WH-questions) indicates that at least some WH-words must induce inversion, if SAI takes place in yes-no questions; yet WH-questions do not appear to contain inverted auxiliaries.

Faced with this contradiction, three options are possible. First of all, the child can choose to ignore the non-inversion evidence in contracted WH-questions, and assume that SAI does indeed take place with at least some WH-phrases. This path will eventually (presumably, through the evidence of uncontracted inverted auxiliaries in WH-questions) lead to the adult grammar.

Second, the child can choose to ignore the inversion evidence in yes-no questions; this choice will lead to the non-inversion stage as exemplified in the grammar of K., where inversion takes place in neither WH- nor yes-no questions.

Third, the child can choose to allow the PLD to temporarily over-ride UG itself, in the form of the implicational relation between inversion in yes-no and WH-questions. The resulting grammar will permit inversion in yes-no questions but not in WH-questions, as in the non-inverting children first described by Bellugi.

Now, of these three options, only the first allows a permanent resolution of the contradiction between the PLD and UG, since further PLD will show the child that WH-questions do indeed induce SAI. On the other hand, both options two and three are at best temporary solutions: one involves ignoring the PLD, the other principles of UG, and no further evidence can resolve the contradiction. In this sense, both of these "non-inverting" options are unstable; as such, we expect them to decay over time, eventually giving way to the adult system when the segmentation errors which led to non-inversion in the first place are corrected.

Thus we see that the non-inversion "stage" is <u>inherently</u> temporary; rather than being a possible adult grammar, it is a transitional attempt to resolve a contradiction between the PLD and UG, which disappears as soon as the child grasps the connection between cliticized auxiliaries in WH-questions and

inverted auxiliaries in yes-no questions.

Note also that we now account for the "missing" non-inversion pattern involving non-inversion in yes-no but not WH-questions. On our account, the child learning SAI could simply never arrive at such a grammar, since it would fly in the face of both UG and the PLD; the crucial point about the two possible non-inversion patterns is that they involve apparent contradiction between UG and the PLD.

We have now accounted for the third of our seven characteristic properties of SAI development. In 5.8, we will expand on our account of non-inversion in SAI development, making the prediction that all "stages" in the acquisition of syntax involve the same type of UG-PLD contradiction. Now, however, let us move on to the fourth and fifth characteristic properties of SAI acquisition, involving "lexical" effects in the development of SAI in WH-questions, and the overgeneralization of how come.

5.74 Closed-class Learning and the Acquisition of SAI

In 5.72, we pointed out that the major differences between the acquisition of INFL2 and SAI could be accounted for by the fact that the former involved much less closed-class learning than the latter. In this sub-section we will review the characteristic properties of closed-class learning, as embodied in the Parallel Distributed Processing model we have adopted from McLelland and Rumelhart (1986), and demonstrate that the

"lexical" effects which are so typical of SAI development fall out naturally from the way in which this learning mechanism operates.

The PDP learning mechanism as described by McLelland and Rumelhart (see Chapter 1) has the following characteristics:

(i) It is sensitive to input-frequency, in that frequency affects the relative "weights" given to particular input-output association lines.

(ii) It has a built-in stochastic element, in the form of a logistic probability function.

(iii) It is "interactive", in that for any given input, all output associations are activated.

McLelland and Rumelhart (1986, Chapter 18) show that, given certain more or less plausible decisions abou data presentation, this model can be used to simulate the "U-shaped learning curve" typical of the acquisition of inflectional morphology; moreover, it can mimic certain other salient properties of morphological learning: namely, token-variability and lexically-based overgeneralization.

Now, we have argued that such properties are typical not only of the learning of inflectional morphology but of the learning of closed-class categories in general. In that case, given that the acquisition of inversion in English involves

closed-class learning, we should expect to find these properties showing up in the developmental patterns which characterize the learning of SAI.

Of course, this is exactly what does happen: as we have seen throughout this chapter (see in particular the summary in 5.44), the learning of SAI in WH-questions is characterized by "lexical" effects, type- and token- variability, and the slow, uneven development characteristic of the acquisition not only of inflectional morphology but of closed-class items in general.

Moreover, inversion with <u>how come</u> is an obvious case of lexical overgeneralization, and shows particularly clearly that the U-shaped learning curve typical of inflectional acquisition also characterizes purely <u>syntactic</u> closed-class learning: initial undergeneralization of SAI (i.e, non-inversion) is followed by item-by-item generalization, and then by limited over-generalization.

This pattern also surfaces in the limited overgeneralization of SAI to embedded clauses (Pinker and Stromswold 1986); once again, the U-shaped learning curve appears to characterize a purely syntactic process.

By adopting the PDP learning mechanism, we also avoid several problems attendant on more conventional rule-based accounts of SAI acquisition. In particular, since "overgeneralization" in this model is not due to hypothesis-formation, but rather to the effect of input frequency on the PDP network at a given point in development, we completely

avoid the necessity for "retreat", with attendant problems involving the use of negative evidence; as soon as the relative weights on the association lines in the PDP pattern associator change as a result of more input, the "overgeneralization" will disappear of its own accord.

Moreover, by accounting for "lexical" learning in terms of the learning mechanism, we considerably simplify and generalize the grammar itself; SAI, for example, can be treated as a completely general syntactic process, even though the effects of closed-class learning surface in the adult language as lexical irregularities.

I have now accounted for characteristics (iv) and (v) in 5.70 above. This leaves two residual properties of SAI development in English: inversion in embedded clauses (property (vi)), and the inhibitory efect of negation on SAI (property (vii)). These will be the subject-matter for the next and final sub-section.

5.75 Residual Problems

I have titled this sub-section "residual problems", because frankly I am much less sure of how to account for properties (vi) and (vii) in 5.70 than for the others we have been investigating in this section. Our discussion will therefore be somewhat brief and speculative; a more explanatory account must await future research.

First of all, then, let us turn to the question of inversion

in embedded clauses. As noted above, embedded-inversion errors typically occur after the non-inversion stage, and follow the U-shaped learning pattern characteristic of closed-class learning. Moreover, in this respect SAI development contrasts strongly with INFL2 development; embedded-inversion errors apparently never occur in the latter.

What could account for the difference ? Recall that in the theoretical framework we have developed for SAI, inversion in subordinate clauses is prevented because embedded WH-phrases are part of a higher AGR-domain, and each AGR-domain has a unique A'-type P-subject (see Chapter 4 for details). Thus, the child who makes embedded-inversion errors must be treating WH-phrases as part of a separate AGR-domain.

Now, we made the observation in 5.64 above that at about the same time that children are learning SAI, they consistently omit the complementizer that both in spontaneous speech and in imitations (see Phinney 1981 for similar findings). This suggests that such children might not have yet have fully distinguished between S (=IP) and S' (=CP); if not, it is possible that they would allow WH-phrases to adjoin to S, a position outside of a higher AGR-domain and therefore liable to induce inversion. In other words, at this stage, children learning SAI would be treating English like Spanish, and thus allowing inversion in embedded clauses.

While this explanation accounts for embedded-inversion errors, it should be pointed out that it raises other questions which are beyond the scope of this thesis. To start with, I

currently have no explanation for why children learning English should consistently omit complementizers. One possibility is that at the stage in question, the only PLD available to the child is of degree-0 (i.e., available in matrix clauses); however, though children initially omit the complementizer if or substitute a WH-word (as we saw in our investigation of K.'s syntactic development) by the time that embedded inversion errors occur, this is no longer the case; therefore at least some degree-1 information must be available at this stage. Another possibility is suggested by the work of Nishigauchi & Roeper (1987), who point out that even such closed-class categories as complementizers are initially acquired as referential elements (e.g., deictic that, benefactive for) and only later acquire non-referential syntactic functions for the language-learner. Thus it is possible that at the stage in question children are still in the "boot-strapping" stage of complementizer acquisition and have not reached the purely syntactic stage. It would be interesting to get some cross-linguistic data on children's acquisition of complementizers; currently I know of none. Clearly, this is an area where more research is necessary.

Finally, we turn to an even more confusing issue: the influence of negation on SAI. We have seen that though data is fairly sparse, Bellugi's original claim that negation inhibited inversion in WH-questions has stood the test of time better than most of her other claims; however, noone has yet produced a satisfactory explanation for her findings.

As we have pointed out several times in this chapter, the problem is that while no syntactic explanation is currently available to explain the inhibitory effect of negation, no non-syntactic explanation is capable of accounting for its specifically syntactic effects. It is not that children have difficulty interpreting or producing negative questions per se, but that they have difficulty producing inverted questions; and after all, inversion is forced by strictly syntactic conditions.

It thus appears that what is missing is the right type of syntactic framework. Note that the negative particle not has other puzzling syntactic properties; it is the only element which blocks [I]/[V] association in English, for example, as noted in Chapter 3. This suggests that it is some idiosyncratic syntactic property of not (or its contracted variants) which causes the inhibitory effect on inversion. One possibility is that there is an interaction between the scope-taking properties of not and question operators; in what follows, we will provide a brief and somewhat speculative account of how an explanation for the inhibitory effect of negation on SAI might be constructed on this basis.

First of all, we will adopt May's (1985) government-based account of scope relations. In essence, May claims that an operator takes scope over its maximal government domain. This means that if two operators govern each other, either may take scope over the other. Thus in a sentence such as (75a.) below, there are two readings, the "distributed" and "collective"

readings, paraphrased in (75b.) and (75c.). On the first reading, the universal quiantifier everyone takes scope over the existential quantifier someone; on the second, the reverse scope-relation holds. However, in both cases the LF rule of Q(uantifier) R(aising) has adjoined the two quantifiers to S; since they govern each other, either reading is possible:

(75)a. Everyone loves someone

b. For all x, there is some y such that x loves y.

c. There is some y such that for all x, x loves y.

Now, suppose we were to claim that children at a certain stage in development do not permit not to take scope over a question operator. Since (following Jackendoff 1972) it is generally assumed that the scope of not is determined by its S-structure position (i.e, it is not subject to QR) this condition will be satisfied unless not is contracted onto an auxiliary and adjoined to S; in that case, by May's government condition on scope, not and the question operator will govern each other and therefore may take scope over each other. Hence, there will be no inversion in negative questions.

Note that this explanation will only work if children do not have the option of inverting the auxiliary and leaving not behind; thus we do not expect children who do not invert in negative questions to produce sentences such as those in (76):

(76)a. Can you not see me ?

b. Do you not want to go to bed ?

c. Will you not come home ?

As far as I can tell, such structures are never in fact produced by non-inverting children; indeed, they have a distinctly "literary" flavour.

While mechanically adequate, this explanation leaves a lot of questions unanswered; in particular, it is quite unclear why the condition on scope proposed above should hold for non-inverting children but not for adults. Once again, it is beyond the scope (excuse the pun) of this thesis to enter into these questions in more detail; clearly, there is a great deal of interesting research to be done here.

This completes my account of the acquisition of SAI. While I have not satisfactorily dealt with all the intricate and puzzling aspects of the learning of inversion, I think I can claim that the explanation developed here allows an interesting and generally successful account of the most salient aspects of SAI acquisition, and provides indirect support both for the syntactic framework for SAI-type inversion developed in Chapter 4 and for the acquisition model I have been developing throughout this thesis.

5.8 SAI Acquisition, Parameter-Setting, and the Two-Tiered Model of Syntactic Development

5.80 Introduction

In the final section of this chapter, we will return to the question of the relation between real-time acquisition and syntactic parameter-setting, before producing an explicit formulation of the "two-tiered" model of syntactic acquisition that has emerged from our investigation of SAI development.

5.81 Linear Parameter-Setting and the Acquisition of SAI

First of all, let us return to the predictions which we made for SAI acquisition based on the parametric model developed in Chapter 4, and repeated below in (77):

(77)a. Children learning English will pass through two stages in the acquisition of SAI, one involving no inversion, the other involving an "INFL2"-type stage.

b. Children learning English will never make NCC-type errors.

c. Children learning English will not apply SAI to exclamatives unless they do so with interrogatives.

d. Children learning English will only produce declarative non-canonical clausal types if they produce interrogative non-canonical clausal types, and interrogative non-canonical clausal types if they also

produce exclamative non-canonical clausal types.

Of the four predictions in (77), (c.) and (d.) are largely irrelevant, since children learning English at the stages we have been studying produce few if any exclamative WH-phrases (though they do appear to be able to imitate them correctly, as evidenced by K.: see 5.33 above). This leaves (a.) and (b.). (b.) appears to be correct: as we have seen, children acquiring English do not at first employ complementizers, yet do not invert when complementizers are absent. On the other hand, (a.), the most important prediction made by a linear-parametric version of the model of SAI-type inversion preseented in Chapter 4, is quite clearly wrong; there is no "INFL2 stage" in the acquisition of SAI.

Faced with this, we have two options: we must either abandon the analysis presented in Chapter 4 or the linear-parametric model itself. By now, it should be obvious which option I intend to take; evidence has been mounting against the linear-parametric model throughout this dissertation, and it is time to replace it with a more plausible model of syntactic acquisition. However, before doing so, it will be helpful once more to examine the inherent weaknesses and strengths of the linear parameter setting, since we wish to preserve the latter without at the same time incorporating the former.

First of all, let us examine the major advantages of the linear parametric model. It turns out that these advantages are almost exclusively meta-theoretical: in particular, linear

parameter setting creates an important connection between cross-linguistic and acquisitional markedness and the theory of learnability, and thus holds out the promise of a reconciliation between syntactic theory, "logical" language acquisition, and real-time language acquisition; this is surely what has made it so attractive to those of us who are committed to reestablishing the connection between these three types of linguistic endeavour.

On the other hand, linear parameter-setting has met with very limited empirical success. Of course, it is always possible that this is simply because we have failed to find the right parameters; however, our investigation of a number of applications of the linear-parametric model to data from language acquisition seems to indicate that the problems are rather more profound. In particular, we have identified the following inherent weaknesses:

(i) The linear-parametric model predicts universal stages in language acquisition which fail to materialize.

(ii) The linear-parametric model cannot deal with inter-subject variation, i.e., children who take different routes to the same adult grammar.

(iii) The linear-parametric model cannot handle mis-matches between cross-linguistic markedness and acquisitional markedness.

(iv) The linear-parametric model encounters problems with the (non-) availability of triggering evidence: why should evidence which has always been accessible only become available to the child at a particular point in grammatical development ?

Bearing both the advantages and disadvantages of the linear-parameter setting model in mind, let us then turn to a model of syntactic acquisition which I hope will offer a plausible alternative: the two-tiered model of parameter setting.

5.83 A Two-Tiered Model of Syntactic Acquisition

In this section, I will attempt to outline a model of syntactic development which retains the conceptual elegance of the linear-parametric model without at the same time also retaining its most conspicuous drawbacks. The central idea behind this model is that syntactic development takes place in two discrete stages, which I will call syntactic recognition and syntactic instantiation: hence the "two-tiered" model. Let us briefly describe each of these types of learning.

The first type of learning, syntactic recognition, characterizes much early "pre-productive" acquisition. Such acquisition occurs rapidly, and without error; it is not tied to particular morphological forms or lexical items, but takes

place in a completely general fashion. It appears likely that some of the more fundamental parameters of Case and θ-theory are set during the recognition period (for example, the deep and surface ergative parameters, which are prerequisites for the acquisition of phrase-structure in the model adopted here - see Chapter 2 for details).

On the other hand, syntactic instantiation, characteristic of "productive" learning, is closely tied to inflectional morphology, sensitive to the properties of specific (closed-class) lexical items, and subject to error in acquisition. It is this type of learning which we have been investigating in our examination of the acquisition of the auxiliary system in English.

Now, by dividing syntactic acquisition into these two very different types of learning, we can begin to account for the "peculiar mixture of general and specific learning" which characterizes the development of SAI and indeed the acquisition of syntax in general. We also free the theory of parameter-setting from the trammels of lexical specificity by obviating the need to tie general parameters to particular lexical items; instead, parameters are set in full generality at the recognition stage but instantiated by particular lexical items at the instantiation stage. In this way we restore the central insight of the parameter-setting model, that a few general parameters can account for a wide range of syntactic variation, without at the same time sacrificing the real-time acquisition data (which, as we have seen, is characterized by a high degree

of lexical specificity).

Aside from these conceptual advantages, the two-tiered model also provides us with a much more plausible account of real-time first language acquisition. In particular, we are now in a position to locate the source of syntactic errors in acquisition, and to understand why only a very small subset of possible errors actually occur. To be precise, I would like to claim that all syntactic errors in acquisition are instantiation errors; that is, there are no recognition errors, or, in other words, no parameter mis-settings. This is a radically different view from conventional linear-parametric models, which envisage a series of parameter-settings, corresponding to "stages" in language acquisition - but as we have seen, there is little or no evidence for such stages in real-time language development.

In this sense, parameter-setting in the two-tiered model - i.e., syntactic recognition - is more powerful and more efficient than in the linear-parametric model. However, the two-tiered learner, unlike the linear-parametric learner, is also faced with the problem of instantiation, and it is here that the potential for syntactic error lies. For example, the child learning SAI knows the fundamental properties of the rule (i.e., that it is forced by conditions on A'-type predication, and thus involves AGR and an A'-type P-subject) but has not correctly identified the AGR-morphology which will enable him or her to apply it correctly: hence, the non-inversion stage. This leads us to the following general hypothesis:

(78) All production errors in the acquisition of syntax are inflectionally driven.

In other words, we predict that all syntactic production errors in acquisition will be linked to misanalyses of inflectional morphology. We have shown that this is indeed the case in the acquisition of SAI, where the non-inverting child reanalyzes contracted auxiliaries as AGR-clitics; likewise, the auxiliary overmarking errors discussed in Chapter 3 are connected to irregular verb morphology. Note that we still allow for the possibility that errors of interpretation have different sources; this seems reasonable, in view of the large number of independent factors (e.g., pragmatics) influencing syntactic comprehension.

The hypothesis in (78) has an additional advantage, in that it enables us to re-establish a direct connection between errors in language acquisition and language change. It is well-known that syntactic change is inflectionally-driven - see for example Lightfoot (1979) and Roberts (1985) on the development of the English modal system - and through (78) we can directly incorporate this generalization in our two-tiered theory of language acquisition: any syntactic error in acquisition is a possible direction for language change. On the other hand, it is less clear how the linear-parametric model could account for this connection without additional stipulation, since intermediate parameter-settings are not inflectionally

determined, and therefore do not necesssarily represent possible directions for change.

This in turn leads us to a rather different conception of syntactic markedness. What is marked cross-linguistically (for example, the "ergative" setting of the "deep" ergative parameter, which characterizes only a tiny subset of the world's languages) is not necessarily marked acquisitionally, or, by hypothesis, liable to syntactic change. Thus, although Greenland Eskimo appears to be - according to current opinion - the only language to set the deep ergative parameter in the marked direction, this does not necessarily mean that the Greenland Eskimo system is unstable and liable to change to a less (cross-linguistically) marked system, as long as there is no morphological opacity which might lead to inflectional misanalysis. In other words, a system may be marked in a cross-linguistic sense but quite stable from the point of view of acquisitional markedness (as long as there are no problems of syntactic instantiation), since such a system is (by definition) syntactically recognizable.

Again, this leads to rather different predictions from the linear-parametric model, which makes a direct connection between cross-linguistic and acquisitional markedness. Within the linear-parametric model, we should always expect languages to change in the direction of a less marked system (i.e., an earlier parameter setting) and, indeed, it becomes quite difficult to explain how grammars become more cross-linguistically marked; in the two-tiered system, on the other

hand, inflectional opacity can drive a language in either a less or a more marked direction, though presumably syntactic recognition will favour the former, ceteris paribus.

To summarize, the two-tiered system, incorporating two distinct types of learning - syntactic recognition and syntactic instantiation - has several advantages over the linear-parametric model, in at least the following domains:

(a) In the domain of syntactic parameter-setting: the two-tiered model allows us to disentangle the theory of parameter-setting from lexical effects and restore it to full generality.

(b) In the domain of real-time language acquisition: the two-tiered model affords us a more accurate set of predictions concerning the nature and scope of syntactic errors in language acquisition.

(c) In the domain of the relation between language acquisition and language change: the two-tiered model reestablishes a direct connection between acquisitional markedness and the direction of syntactic change.

(d) In the domain of syntactic markedness. The two-tiered model does not commit us to the linear-parametric hypothesis that whatever is cross-linguisticaly marked is necessarily acquisitionally marked, and thus avoids the paradox of

explaining how marked systems ever arise when the direction of change should always be in the direction of a less marked system.

This has been a somewhat cursory overview of a theory of language acquisition which I believe shows a good deal of promise for future research, though obviously at present it is in its infancy. I do believe, however, that even from the sketch presented above it can be seen that the two-tiered model has several major empirical and conceptual advantages over the linear-parametric model, and thus has the potential to provide a viable alternative as a model of syntactic acquisition.

CHAPTER SIX

CONCLUSION: ON THE RELATION BETWEEN THE THEORY OF GRAMMAR AND THE STUDY OF LANGUAGE ACQUISITION

6.0 Introduction

At the outset of this thesis, we discussed the peculiarly paradoxical situation which the study of language acquisition currently finds itself in. On the one hand, the theory of syntax is becoming - at least under the Chomskyan interpretation - more and more explicitly a theory of a certain component of language acquisition, i.e, Universal Grammar; on the other, real-time language acquisition data still appears for the most part to to be of little relevance to syntactic theory-construction. At best, evidence from language acquisition is drafted in as a form of novel - and somewhat exotic - "external evidence"; at worst, it is ignored completely. In other words, the bridge constructed by learnability between language acquisition and syntactic theory still does not bear two-way traffic, pace Wexler and Culicover (1980).

In this (mercifully) short concluding chapter, we will be arguing on the basis of evidence from the rest of the thesis that a good deal more construction work must go on go on before the bridge is safe enough to support heavy traffic, and that

construction must go on from both sides of the river. In 6.1, we will examine the problem from the perspective of the responsibility of syntactic theory to language acquisition; in 6.2, we will go on to investigate the view from the other side of the river, discussing the responsibility of language acquisition to syntactic theory.

6.1 The Responsibility of Syntactic Theory to Language acquisition

6.10 Introduction

While it is true to say that data from real-time language acquisition has had little impact on syntactic theory, it is certainly not the case that the issue of language acquisition per se has been ignored by syntacticians. In fact, one of the most influential theoretical innovations in the recent history of syntax has been the perspective provided by what has become known as the "logical" theory of language acquisition, developed originally in the work of Grimshaw (1979), Baker (1979), Lightfoot (1979), and Wexler and Culicover (1980), amongst others. In this section, we will be examining ways in which this "logical" theory of language acquisition can be extended to further restrict the theory of syntax and bring the theory closer to a "realistic" view of language acquisition.

6.11 From "Instantaneous" to "Logical" Language Acquisition

First of all, I would like to point out that the "logical" theory of language acquisition is a misnomer (after all, noone is suggesting a rival "illogical" theory !); rather, what is generally meant by logical acquisition is in fact idealized acquisition. I would like to argue, however, that this idealized theory of language acquisition is part of a logical development in the relationship between syntactic theory and real-time language acquisition, a development which began with the "instantaneous" theory of Chomsky (1965) and will end - I hope - in the full integration of the two disciplines. Let us briefly retrace the first steps in this development.

As is well-known, Chomsky (1965) introduced both the idea that a syntactic theory should be construed at a certain level of abstraction as a theory of language acquisition, and the level of abstraction itself, in the form of the idealization of instantaneous acquisition. Thus the original agenda of the "syntax-as-acquisition" research paradigm already contained the seeds of the paradox that characterizes the field twenty years later; though in the instantaneous model acquisition is of great meta-theoretical impact, it is conceived of in such an abstract way that real-time studies of language learning can have no influence on the formulation of grammars.

Viewed from the perspective of language acquisition, much of the history of generative syntax in the last twenty years can be seen as involving a gradual shift from the completely idealized instantaneous model towards a more plausible, if

still abstract, model of language learning. Pinker (1979) identifies the following six empirical conditions on a theory of language learning:

(1)a. Learnability
   b. Equipotentiality
   c. Time
   d. Input
   e. Developmental
   f. Cognitive

Of these six conditions, (a), (b), and (d) now play a prominent part in syntactic theory. At the risk of sending the reader to sleep, we will briefly detail their impact here, before going on to discuss ways in which the theory of language acquisition can further influence the development of a cognitively plausible model of syntax.

The learnability condition is the subject of learnability theory, which applies the axioms and theorems of mathematical models of language-learning to natural language. Learnability theory, of course, is one of the most prominent and influential outgrowths of the grammar-as-acquisition paradigm (see Chapter 1); learnability considerations have been behind the dominant trend towards "restrictiveness" in grammar construction, which in turn has led to the two major "learnability positions" now adopted by generative syntacticans. On the one hand, there are those who believe that a learnable theory of syntax can best be constructed by restricting the weak generative power of permissible human grammars; this is the position taken by advocates of Generalized Phrase Structure Grammar, for example

(see Gazdar, Klein, Pullum and Sag 1985). On the other hand, there are those who claim that by imposing a finite limit on the total number of grammars, the learnability condition can be met without reference to classes of languages as defined by the criterion of generative power (this is the Chomskyan position). In fact, the latter view leads rather inevitably to a still more radical notion of restrictiveness: the claim that UG consists of a single grammar, with a finite set of parameters; this is the position taken by contemporary parameter-setting theories of syntax, and the one which we have adopted in this thesis.

Let us then turn to the equipotentiality condition, which is meant to capture the fact that a child is capable of learning any possible human grammar, and not just, say, the grammars of Mohawk and Albanian. In recent years, the problem of equipotentiality has inspired a renewed interest in comparative linguistics, and the development of the parametric model of grammar, which has enabled syntacticians to come to grips with cross-linguistic variation within the framework of Universal Grammar, has led to the development of an increasingly sophisticated and interesting research paradigm. Parts of Chapters 2, 3, and all of Chapter 4 are cast within this paradigm.

Finally, we turn to the input condition. The principal influence of input conditions on syntactic theory has been the assumption that children receive no negative evidence in the course of learning a natural language. This assumption

profoundly affects language learnability (see Chapter 1 for comments) and has led to the development of such influential constraints on language acquisition as the Subset Principle of Berwick (1982). In terms of the theory of syntax, it is now common practice to organize a parametrized grammar so that each value of a parameter is "de-learnable" using positive evidence only; this is of course directly reflected in the linear-parametric model of real-time language acquisition, exemplified by the work of Hyams (1983) - see Chapters 1 and 2 - and Klein (1982) - see Chapter 5.

Thus we see that in fact the influence of language acquisition on syntactic theory-construction has greatly increased since the days of "instantaneous acquisition"; of the six conditions in (1) above, three have been directly or indirectly incorporated into grammatical theory.

Now, the question arises as to the syntactic status of the other three conditions: the time, developmental, and cognitive conditions. Which, if any, should be incorporated next ?

The most promising of the three is surely the developmental condition; indeed, much of this thesis has been devoted to exploring the question of how developmental sequences might affect grammars, and vice-versa. In the next sub-section, we will make some suggestions as to how the developmental condition should best be instantiated in the theory of syntax, in the light of the investigation we have conducted in the last five chapters.

6.12 From "Logical" to "Plausible" Language Acquisition

One initially attractive way of incorporating the developmental condition is by treating acquisition sequences as sequences of possible grammars, ordered by UG with respect to some criterion of markedness. This is of course the approach taken by the linear-parametric model of syntax, incorporating what we might call a markedness correspondance between cross-linguistic variation and developmental sequences; the more marked a parameter setting, the later it will be hypothesized by the child and the less likely it will be instantiated cross-linguistically. As we have seen, the linear-parametric model makes strong predictions about what we should find in the acquisition data: universal developmental stages, each corresponding to a possible adult grammar, ordered by the logic of learnability and the theory of parametric markedness.

On the basis of our investigation, I think we can say fairly safely that these predictions are not borne out by the data. We have examined three areas of syntactic acquisition, at least two of which have been subjected to linear-parametric treatments (those of Klein 1982 and Hyams 1983); in no case have we found evidence for universal stages, nor for the markedness correspondance predicted to exist by the linear-parametric model.

In some ways, this is a shame; the linear-parametric model has an undeniable conceptual neatness, especially in the often confusing and confused field of real-time language acquisition.

On the other hand, its very neatness has always appeared too good to be true, especially to veteran child language researchers who have seen several syntactic theories arrive amidst great pomp and ceremony, only to be ignominiously dumped a couple of years later.

What went wrong ? As we have argued in Chapter 5, the answer seems to be that the linear-parametric model is based on far too simple a view of the way in which syntactic competence (in the form of UG) is embedded in a particular type of performance (language acquisition). In 5.7, we presented an alternative model, based on two very distinct types of learning, which claimed that language acquisition was at the same time easier and more difficult than predicted by linear parameter setting: easier, since major parameters appear to be set - correctly - very early in acquisition; more difficult, because the details of the morphological realization of those parameters must be worked out slowly and painfully over several years.

Where, then, does this leave the relationship between the theory of syntax and that of language acquisition ? Given the failure of the linear-parametric model, should we abandon - at least temporarily - the attempt to incorporate developmental criteria into syntactic theory ?

I believe that the answer to this question is no. The problem seems to me to lie not so much in the general project of attempting to incorporate acquisition criteria into syntactic theory, as in the specific assumptions made by

the proponents of the linear-parametric model. In fact, it seems to me that the linear-parametric approach is much too ambitious, given our currrent state of knowledge; at the risk of over-burdening our bridge metaphor, linear-parameter setting is an attempt to jump the river - a kind of Eivel Kneivel approach to the relationship between syntactic theory and real-time language acquisition.

How, then, ought we to incorporate developmental criteria into the theory of syntax, so as to avoid the fate of the linear-parametric model ? One way of doing so is to begin to develop a notion of what Chomsky (1981) has referrd to as epistemological priority: theory construction should be based on principles and concepts that could plausibly be available to the pre-linguistic child.

In Chapter 2, I argued that an approach based on epistemological priority might represent the next logical step in the rapprochement (or perhaps, simply approchement) of syntactic theory with language acquisition. I pointed out that applying this criterion to one contemporary syntactic theory (the GB model) had rather profound consequences for the structure of the grammar, and went on to argue that those consequences were desirable in both purely syntactic terms and in terms of the relation between syntactic theory and the theory of language acquisition. By developing an explicit theory of phrase-structure acquisition based on the interaction of "epistemologically primitive" components of the theory, I hope to have shown that GB theory can in fact be mapped into a

plausible model of early language acquisition, which simulates reasonably accurately the course of real-time language acquisition.

Thus it seems to me that the criterion of acquisitional plausibility, as exemplified by the notion of epistemological priority, represents the most promising way of incorporating developmental considerations into syntactic theory, given our current state of knowledge.

Let us now cross the river and enter the wilder territory of real-time language acquisition.

6.2 The Responsibility of Language Acquisition to Syntactic Theory

6.20 Introduction

In this section, we will be examining the place of syntactic theory in the field of real-time language acquisition. We will not try to justify the claim that any even vaguely sensible theory of language acquisition must incorporate a theory of grammar; this has been done many times before, and those who aren't convinced by now will probably never be convinced. Since a language is by definition the product of a grammar, an account of language acquisition without a grammar is like a baby born without a brain; it may survive for a while on life-support systems, but it doesn't have a future.

In fact we will be arguing that the problem with most attempts to incorporate syntactic theory into real-time language acquisition is not that they have failed to take grammars into account, but that they have failed to take learning into account. It will follow that the chief responsibility of a theory of language acquisition is not the incorporation of grammatical theories per se, but the delimitation of the domain of those theories.

In the following sections we will argue that the focus of the study of language acquisition should shift to those aspects of development which are <u>not</u> accounted for by syntactic theory, but nonetheless affect both language learning and the form of the adult language.

At the very beginning of this thesis, we presented four necessary components of a theory of language acquisition:

(2)a. The nature of the adult grammar (S_s)
   b. The nature of the initial state (S_o)
   c. The nature of the mediating evidence (PLD)
   d. The nature of the "learning" mechanism that effects the transition(s) from S_o to S_s.

Of these components, (c) and (d) are quite clearly within the domain of the theory of language learning, not of the theory of grammar. This is not to say, of course, that they cannot influence the way grammatical theories are constructed; as we have seen, the "no-negative-evidence" dictum has had enormous

influence on the theory of syntax; the point is, the fact that children do not attend to overt corrections is an empirical discovery about human language acquisition, and as such, can only come from the study of real-time language development.

In fact, I believe that this type of empirical discovery represents the most tangible contribution of the study of real-time language acquisition to the theory of grammar. In a sense, this contribution is indirect; rather than finding ways of mapping grammars directly into acquisition sequences, it involves discovering the conditions under which language learning takes place, and tracing the influence of those conditions on the form of the resulting grammars. In this way, we can begin to sort out those aspects of human language that are determined by the process of real-time language-learning, and those that are directly determined by UG; through this delimitation procedure, we begin to get a clearer picture of what is language-specific, and must be part of our innate linguistic competence, and what can be accounted for by more general conditions on human learning.

In the next two sections we will very briefly review the two most important ways in which the study of real-time language acquisition can contribute to syntactic theory: the study of input conditions, and the study of learning processes, corresponding to 2(c) and 2(d) above, respectively.

6.21 Input conditions

We have already pointed out that one restriction on input (the "no-negative-evidence" dictum) has had a profound effect on theory construction; another (that based on "degrees of embedding") is also beginning to be taken seriously as a restriction on syntactic theories (see Wexler and Culicover 1980, Culicover and Wilkins 1984).

However, in spite of the importance of the issue, the fact is we still know next to nothing about what kind of evidence the child uses when, and why (see Newport, Gleitman, and Gleitman 1977 for one shining exception to this gloomy generalization). As we have seen, studies within the framework of the linear-parametric model are particularly prone to problems caused by the child's apparently ignoring readily available PLD until a certain point in development is reached, when it suddenly becomes crucial triggering evidence. An example discussed extensively in the last chapter is the issue of why fronted auxiliaries in yes-no questions, which are salient enough to help children learn auxiliaries in medial positions, are not salient enough to help them learn inversion in WH-questions during the non-inversion stage.

Thus, while it is clear that the relationship between the child's developing grammar and the PLD is indirect, we have no idea how exactly the relationship is mediated. One plausible approach, which we adopted in Chapter 5, is that the child's grammar itself acts as a filter on the input; but then we were faced with the problem of how the child "breaks out" of an intermediate grammar. We suggested that the solution to

this problem might lie in the inherent "instability" of transitional grammars in acquisition, but at this point, such a solution is speculative: clearly, this is an area which could handle a good deal more research.

In conclusion, the study of input conditions has proven particularly fruitful in the relation of real-time language acquisition to syntactic theory, and, I believe, has a lot still to offer; certainly, we know how to ask the right questions, even if we don't know to set about answering them at this stage.

6.22 Learning Mechanisms

One of the major hypotheses of this thesis has been the proposal that much of the work currrently done by syntactic theories in language acquisition can be taken over much more successfully by an appropriate learning mechanism. We have shown, in particular, that a whole set of properties associated with the learning of closed-class syntactic items - inflections, auxiliary verbs, WH-phrases, and so forth - can be explained naturally and without complicating the theory of grammar by adopting a Parallel Distributed Processing Model of the type described by McClelland and Rumelhart (1986). These properties include not only those which can in principle be handled by syntactic theory ("lexical" learning, undergeneralization), but also those which cannot (token variability, lexically-based overgeneralization and "retreat",

and the U-shaped learning curve). As we have seen, such properties characterize not only morphological learning, but the learning of purely "syntactic" rules such as SAI.

It should be emphasized once again that we do not believe that a PDP model can replace a grammar; rather, given an appropriate division of labour, it can considerably simplify the grammar by relieving it of the task of dealing with lexical irregularity. Moreover, the fact that it is closed-class learning which seems to be mediated by a PDP mechanism is itself of considerable theoretical interest, tending to support recent neurolinguistic findings that there is a fundamental distinction between open- and closed-class elements in natural language, and to reinforce the purely syntactic distinction captured by Emonds (1985) in the form of the principle of Unique Syntactic Behaviour and the Designation Convention, discussed in Chapters 3 and 4.

In conclusion, I believe that a more articulated theory of learning mechanisms can make it much easier for us to delimit the role of the grammar in language acquisition. This dissertation has only begun to explore the implications of this approach, but I believe the results are encouraging enough to merit further investigation, both from the standpoint of real-time language acquisition and that of syntactic theory.

6.3 Conclusion

We have now peered across the chasm separating syntactic theory

from the study of real-time language acquisition from both sides. I have argued, as I have argued throughout this thesis, that if a bridge is to be constructed linking the two fields, it will have to be built by practicioners of both disciplines, from either side of the river; and I have suggested some ways in which the two theories can begin to bridge the gulf.

As an intellectual enterprise, the task of linking theoretical syntax with real-time language acquisition seems to me deeply exciting, though often perilous. If nothing else, I hope to have communicated some of this excitement, and to have shown that both disciplines can benefit from each other's perspective on the central question in current syntactic theory: the question of how children acquire grammars.

REFERENCES

Abbot, B. (1976). "Right Node Raising as a Test for Constituency", Linguistic Inquiry 7: 639-642.

Abney, S. (1985). "Functor Theory and Licensing: Toward an elimination of the Base Component", ms., MIT, Cambridge, MA.

Adams, M. (1984). "Multiple Interrogation in Italian", The Linguistic Review 4: 1-27.

Akmajian, A., R. A. Demers, & R. M. Harnish (1984). Linguistics: an introduction to language and communication (2nd edition), MIT Press, Cambridge, MA.

Akmajian, A., S. M. Steele, & T. Wasow (1979). "The Category AUX in Universal Grammar", Linguistic Inquiry 10: 1-64.

Akmajian, A., & T. Wasow (1975). "The Constituent Structure of the VP and AUX and the position of the Verb BE", Linguistic Analysis 1: 205-246.

Andrews, A. D. (1982). "The Representation of Case in Modern Icelandic", in J. Bresnan, ed.

Aoun, Y. (1985). A Grammar of Anaphora, MIT Press, Cambridge, MA.

Aoun, Y., N. Hornstein & D. Sportiche (1981). "Some Aspects of Wide-scope Quantification", Journal of Linguistic Research 1: 69-95.

Aoun, Y. & D. Sportiche (1983). "On the Formal Theory of Government", The Linguistic Review 2: 211-236.

Aronoff, M. (1976). Word Formation in Generative Grammar, MIT press, Cambridge, MA.

Bach, E. (1976). "An extension of Classical Transformational Grammars", Problems in Linguistic Metatheory, Proceedings of the 1976 Conference at Michigan State University, 183-224.

Baker, C.L. (1970). "Notes on the description of English Questions: the Role of an Abstract Question Morpheme", Foundations of Language 6: 197-219.

Baker, C.L. (1979). "Syntactic Theory and the Projection Problem", Linguistic Inquiry 10: 533-581.

Baker, C.L. & J.J. McCarthy eds., (1981). The Logical Problem of Language Acquisition, MIT Press, Cambridge, MA.

Baker, M. (1985) Incorporation: a Theory of Grammatical Function Changing, unpublished MIT PhD dissertation.

Bar-Hillel, Y. (1953). "A Quasi-arithmetical Notation for Syntactic Description", Language 29: 47-58.

Barwise, J. and R. Cooper (1981). "Generalized Quantifiers and Natural Language", Linguistics and Philosophy 4: 159-219.

Bayer, J. (1984). "COMP in Bavarian Syntax", The Linguistic Review 3: 209-274.

Belletti, A., & L. Rizzi (1981). "The Syntax of ne: Some Theoretical Implications", The Linguistic Review 1: 117-154.

Bellugi, U. (1967). The Acquisition of the System of Negation in Children's Speech, unpublished Harvard PhD dissertation.

Bellugi, U. (1968). "Linguistic Mechanisms Underlying Child Speech", in H. Zale, ed., Proceedings of the Conference on Language and Language Behaviour, Appleton-Century-Crofts, New York.

Bellugi, U. (1971). "Simplification in Children's Language", in R. Huxley and E. Ingram, eds., Language Acquisition: Models and Methods, Academic Press, New York.

Bellugi, U. & R. Brown, eds. (1964). The Acquisition of Language: Report of the Fourth Conference Sponsored by the Committee on Intellective Process Research of the Social Science Research Council, University of Chicago Press, Chicago.

Belvin, R. (1984). Nishga Syntax and the Ergativity Hypothesis, unpublished University of B.C. M.A. dissertation.

Bennis, H. (1983). "A Case of Restructuring", in H. Bennis & W. U. S. van Lessen Kloeke, eds., Linguistics in the Netherlands 1983, Foris publications, Dordrecht.

Bennis, H. & L. Haegeman (1984). "On the Status of Agreement and Relative Clauses in West-Flemish" in W. deGeest & Y. Putseys, eds., Sentential Complementation, Foris publications, Dordrecht.

Berwick, R. (1982). Locality Principles and the Acquisition of Syntactic Knowledge, unpublished MIT PhD dissertation.

Berwick, R. & A. Weinberg (1984). The Grammatical Basis of Linguistic Performance: Language Use and Acquisition, MIT Press, Cambridge, MA.

Besten, H. den. (1977) (published 1983). "On the Interaction of Root Transformations and lexical deletive rules" in W. Abraham, ed., On the Formal Syntax of the Westgermania, John Benjamins, Amsterdam.

Besten, H. den. (1982). "Some Remarks on the Ergative Hypothesis", Groninger Arbeiten zur germanistischen Linguistik 21: 61-81.

Besten, H. den. (1985). "The Ergative Hypothesis and Free Word Order in Dutch and German", in J.Toman, ed.

Bever, T. (1970). "The Cognitive Basis for Linguistic Structures", in J. R. Hayes, ed.

Bloom, L. (1970). Language Development: Form and Function in Emerging Grammars, MIT Press, Cambridge, MA.

Bloom, L. (1973). One Word at a Time:the Use of Single-word Utterances Before Syntax, Mouton, the Hague.

Bloom, L., P. Lightbown, & L. Hood (1975). Structure and Variation in Language Development, Society for Research in Child Development Monographs 40.

Borer, H. (1984). Parametric Syntax: Case Studies in Semitic and Romance Languages, Foris publications, Dordrecht.

Borer, H. (1986). "I-Subjects", Linguistic Inquiry 17: 375-416.

Borer, H. & K. Wexler (1987). "The Maturation of Syntax", in T. Roeper and E. Williams, eds.

Bouchard, D. (1984). On the Content of Empty Categories, Foris publications, Dordrecht.

Bowerman, M. (1973). Early Syntactic Development: a Cross-linguistic Study with Special Reference to Finnish, Cambridge University Press, Cambridge.

Braine, M. (1976). Children's First Word Combinations, Society for Research in Child Development Monographs 41.

Brandi, L. & P. Cordin (1981). "On Clitics and Inflection in Some Italian Dialects", ms., Scuola Normale Superiore di Pisa.

Bresnan, J., ed. (1982). The Mental Representation of Grammatical Relations, MIT Press, Cambridge, MA.

Brody, M. (1984). "On Contextual Definitions and the Role of Chains", Linguistic Inquiry 15: 355-380.

Brown, R. (1968). "The development of WH-questions in Child Speech", Journal of Verbal Learning and Verbal Behaviour 7: 279-290.

Brown, R. (1973). A First Language: the Early Stages, Harvard University Press, Cambridge, MA.

Brown, R. & C. Fraser (1963). "THe Acquisition of Syntax", in U. Bellugi & R. Brown, eds.

Brown, R. & C. Hanlon (1970). "Derivational Complexity and Order of Acquisition in Child Speech", in J. R. Hayes, ed.

Burzio, L. (1981). Intransitive Verbs and Italian Auxiliaries, unpublished MIT PhD dissertation.

Bybee, J & D. Slobin (1982). "Rules and Schemas in the Development and Use of the English Past Tense", Language 58: 265-289.

Cazden, C. (1972). Child Language and Education, Holt, Rinehart & Winston, New York.

Cedergren, H., & D. Sankoff (1974). "Variable Rules: Performance as a Statistical Reflection of Competence", Language 50: 333-355.

Chomsky, N. (1957). Syntactic Structures, Mouton, The Hague.

Chomsky, N. (1964). "Formal Discussion" in U. Bellugi & R. Brown, eds.

Chomsky, N. (1965). Aspects of the Theory of Syntax, MIT Press, Cambridge, MA.

Chomsky, N. (1970). "Remarks on Nominalization", in R. Jacobs and P. Rosenbaum, eds., Readings in English Transformational Grammar, Ginn & Co., Watham, MA.

Chomsky, N. (1973). "Conditions on Transformations", in S. Anderson and P.Kiparsky, eds., A Festshcrift for Morris Halle, Holt, Rinehart & Winston, New York.

Chomsky, N. (1975). Reflections on Language, Pantheon, New York.

Chomsky, N. (1976). "Conditions on Rules of Grammar", Linguistic Analysis 2: 303-352.

Chomsky, N. (1977). "On WH-movement", in P. Culicover, T. Wasow, & A. Akmajian, eds.

Chomsky, N. (1980a.). "On Binding", Linguistic Inquiry 11: 1-46.

Chomsky, N. (1980b.) Rules and Representations, Columbia University Press, New York.

Chomsky, N. (1981). Lectures on Government and Binding, Foris publications, Dordrecht.

Chomsky, N. (1982). Some Concepts and Consequences of the Theory of Government and Binding, MIT Press, Cambridge, MA.

Chomsky, N. (1986a.) Knowledge of Language: its Nature, Origin, and Use, Praeger, New York.

Chomsky, N. (1986b.) Barriers, MIT Press, Cambridge, MA.

Chomsky, N. & H. Lasnik (1977). "Filters and Control", Linguistic Inquiry 8: 425-504.

Chung, S & J. McCloskey (1983). "On the Interpretation of Certain Island Effects in GPSG", Linguistic Inquiry 14: 704-713.

Clahsen, H. (1982). Spracherwerb in der Kindheit. Eine Untersuchung zur Entwicklung der Syntax bei Kleinkindern. Narr, Tubingen.

Clahsen, H. (1984). "Verb Inflections in German Child Language: the Acquisition Agreement Markings and the Functions they Encode", ms., to appear in Linguistics 25.

Clahsen, H. and K-D. Smolka (1986). "Psycholinguistic Evidence", in H. Haider and M. Prinzhorn, eds.

Contreras, H. (1983). "ECP as an S-Structure Principle", ms., University of Washington, Seattle, WA.

Contreras, H. (1986). "Open and Closed A'-Chains", in P. Coopmans et al., eds., Formal Parameters of Generative Grammar II: Going Romance, Foris publications, Dordrecht.

Cooper, W., & J. Paccia-Cooper (1980). Syntax and Speech, Harvard University Press, Cambridge, MA.

Culicover, P. (1976). Syntax, Academic Press, New York.

Culicover, P., T. Wasow & A. Akmajian, eds. (1977). Formal Syntax, Academic Press, New York.

Culicover, P. & W. Wilkins (1984). Locality in Linguistic Theory, Academic Press, New York.

Culy, C. (1985). "The Complexity of the Vocabulary of Bambara", Linguistics and Philosophy 8: 345-351.

Czepluch, H. (1983). "Case Theory and the Dative Construction", The Linguistic Review 2: 1-38.

Davis, H. (1983). "Restructuring in Child Language", paper given at the Canadian Linguistics Association Conference, Vancouver, B.C.

de Villiers, J. & P. de Villiers (1973). "A Cross-sectional Study of the Acquisition of Grammatical Morphemes", Journal of Psycholinguistic Research 2: 331-341.

de Villiers, J., & P. de Villiers (1974). "Competence and Performance in Child Language: Are Children really Competent to Judge ?" Journal of Child Language 1: 11-22.

Elesseily, N. (1985). Subject Extraction from Embedded Clauses in Standard Arabic, unpublished UBC M.A. dissertation.

Emonds, J. (1976). A Transformational Approach to English Syntax, Academic Press, New York.

Emonds, J. (1978). "The Verbal Complex V'-V in French", Linguistic Inquiry 9: 151-175.

Emonds, J. (1985). A Unified Theory of Syntactic Categories, Foris publications, Dordrecht.

Engdahl, E. (1979). "WH-constructions in Swedish and the Relevance of Subjacency", in Papers from the Eleventh Regional Meeting of the North Eastern Linguistic Society, Ottawa, Ontario.

Erreich, A. (1984). "Learning How to Ask: Patterns of Inversion in Yes-no and WH-questions", Journal of Child Language 11: 579-592.

Erreich, A., V. Valian & J. Winzemer (1980). "Aspects of a Theory of Language Acquisition", Journal of Child Language 7: 157-179.

Evers, A. (1975). The Transformational Cycle in Dutch and German, Indiana University Linguistics Club, Bloomington, Indiana.

Fabb, N. (1983). "Three Squibs on Auxiliaries", MIT Working Papers in Linguistics 5: 104-120.

Fabb, N. (1984). Syntactic Affixation, unpublished MIT PhD dissertation.

Falk, Y. (1984). "The English Auxiliary System", Language 60: 483-509.

Fay, D. (1978). "Transformations as Mental Operations: a Reply

to Kuczaj", Journal of Child Language 5: 143-150.

Fay, D. (1980). "Performing Transformations", in R. Cole, ed., Perception and Production of Fluent Speech, Erlbaum, Hillsdale, New Jersey.

Fee, E. J. (1980).The Relationship Between Mothers' Pronominal Modifications and Children's Acquisition of Pronominal Reference, unpublished University of B.C. M.A. dissertation.

Fletcher, P. (1979). "The Development of the Verb Phrase" in P. Fletcher & M. Garman, eds., Language Acquisition (first edition), Cambridge University Press, Cambridge.

Fletcher, P. & M. Garman, eds. (1986). Language Acquisition (second edition), Cambridge University Press, Cambridge.

Fodor, J., T. Bever & M. Garret (1974). The Psychology of Language: an Introduction to Psycholinguistics and Generative Grammar, McGraw-Hill, New York.

Foss, D. & D. Fay (1975). "Linguistic Theory and Performance Models", in J. Wirth and D. Cohen, eds., Testing Linguistic Hypotheses, Hemisphere Press Inc., New York.

Fraser, C., U. Bellugi, & R. Brown (1963). "Control of Grammar in Imitation, Comprehension, and Production", Journal of Verbal Learning and Verbal Behaviour 2: 121-135.

French, M. (1984). "Markedness and the Acquisition of Pied-piping and Preposition Stranding", McGill University Working Papers in Linguistics 2: 131-145.

Fukui, N. & M. Speas (1986). "Specifiers and Projections", in MIT Working Papers in Linguistics 8.

Gazdar, G. (1982). "Phrase Structure Grammar", in P. Jacobson and G. Pullum, eds.

Gazdar, G., G. Pullum & I. Sag (1982). "Auxiliaries and Related Phenomena in a Restricted Theory of Grammar", Language 58: 591-638.

Gazdar, G., E. Klein, G. Pullum & I. Sag (1985). Generalized Phrase Structure Grammar, Harvard University Press, Cambridge, MA.

George, L. (1980). Analogical Generalizations in Natural Language Syntax, unpublished MIT PhD dissertation.

Givon, T. (1976). "Topic, Pronoun, and Grammatical Agreement", in C. Li, ed.

Gold, E. (1967). "Language Identification in the Limit", Information and Control 16: 447-474.

Goldin-Meadow, S. (1979). "Language Without a Helping Hand", in H. Whitaker & H. A. Whitaker, eds., Studies in Neurolinguistics, Vol. 4, Academic Press, New York.

Goldsmith, J. (1981). "Complementizers and Root Sentences", Linguistic Inquiry 12,: 541-574.

Goodluck, H. & L. Solan (1979). "A Reevaluation of the Basic Operations Hypothesis", Cognition 7: 85-91.

Green, G. (1976). "Main Clause Phenomena in Subordinate Clauses", Language 52: 382-397.

Greenberg, J. (1963). "Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements", in J. Greenberg, ed., Universals of Language, MIT Press, Cambridge, MA.

Grimshaw, J. (1977). English WH-constructions and the Theory of Grammar, unpublished U. Mass. PhD dissertation.

Grimshaw, J. (1979). "Complement Selection and the Lexicon", Linguistic Inquiry 10: 279-326.

Grimshaw, J. (1981). "Form, Function, and the Language Acquisition Device", in C.L. Baker and J. J. McCarthy, eds.

Gruber, J. (1967). "Topicalization in Child Language", Foundations of Language 3: 37-65.

Guilfoyle, E. (1984). "The Acquisition of Tense and the Emergence of Lexical Subjects in Child Grammars of English", McGill University Working Papers in Linguistics 2: 20-30.

deHaan, G. and F. Weerman (1986). "Finiteness and Verb Fronting in Frisian" in H. Haider and M. Prinzhorn, eds.

Haegeman, L. & H. van Reimsdijk (1986). "Verb Projection Raising, Scope, and the Typology of Rules Affecting Verbs", Linguistic Inquiry 17: 417-466.

Haider, H. (1985). "The Case of German", in J. Toman, ed.

Haider, H. (1986a.). "Affect α: A Reply to Lasnik & Saito, 'On the Nature of Proper Government'", Linguistic Inquiry 17: 113-126.

Haider, H. (1986b.). "V-Second in German", in H. Haider & M. Prinzhorn, eds.

Haider, H. & M. Prinzhorn, eds. (1986). Verb Second Phenomena in Germanic Languages, Foris publications, Dordrecht.

Haik, I. (1985). The Syntax of Operators, unpublished MIT PhD dissertation.

Hale, K. (1979). On the Position of Warlpiri in a Typology of the Base, Indiana University Linguistics Club, Bloomington, Indiana.

Hale, K. (1983). "Warlpiri and the Grammar of Non-Configurational Languages", Natural Language and Linguistic Theory 1: 1-43.

Halle, M. (1962). "Phonology in Generative Grammar", Word 18: 54-72.

Halle, M. , J.Bresnan & G. Miller (1978). Linguistic Theory and Psychological Reality, MIT Press, Cambridge, MA.

Hankamer, J., & I. Sag (1976). "Deep and Surface Anaphora", Linguistic Inquiry 7: 391-428.

Hayes, J. R., ed. (1970). Cognition and the Development of Language, Wiley, New York.

Henry, A. (1987). "The Structure of Infinitives", paper given at the 1987 GLOW conference, Venice.

Higginbotham, J. (1985). "A Note on Phrase-Markers", MIT Working Papers in Linguistics 6: 87-101.

Higginbotham, J. (1986). "Elucidation of Meaning, ms., MIT, Boston, MA.

Higginbotham, J. & R. May (1981). "Questions, Quantifiers, and Crossing", The Linguistic Review 1: 41-97.

Hooper, J. & S. Thompson (1973). "On the Applicability of Root Transformations", Linguistic Inquiry 4: 465-498.

Horgan, D. (1978). "The Development of the Full Passive", Journal of Child Language 5: 63-80.

Horning, J. (1969). A Study of Grammatical Inference, Technical Report # CS 139, Computer Science Dept., Stanford University, Stanford, California.

Hornstein, N. & D. Lightfoot, eds. (1981). Explanations in Linguistics, Longmans, London.

Hornstein, N. & A. Weinberg (1981). "Case Theory and Preposition Stranding", Linguistic Inquiry 12: 55-92.

Huang, C.-T. J. (1982). Logical Relations in Chinese and the Theory of Grammar, unpublished MIT PhD dissertation.

Huddleston, R. (1978). "On the Constituent Structure of VP and AUX", Linguistic Analysis 4: 31-59.

Hudson, R. (1976). Arguments for a Non-Transformational Grammar, Chicago University Press, Chicago.

Hurford, J. (1975). "A Child and the English Question Formation Rule", Journal of Child Language 2: 299-301.

Huxley, R. (1970). "The Development of the Correct Use of Subject Personal Pronouns in Two Children", in G. B. Flores d'Arcais and W. M. J. Levelt, eds., Advances in Psycholinguistics, North-Holland, Amsterdam.

Huxley, R. and E. Ingram, eds. (1971). Language Acquisition: Models and Methods, Academic Press, New York.

Hyams, N. (1983). The Acquisition of Parametrized Grammars, unpublished CUNY PhD dissertation.

Hyams, N. (1986). "Core and Peripheral Grammar and the Acquisition of Inflection", paper given at the eleventh Boston University Conference on Language Development, Boston, MA.

Ingram, D. and D. Tyack (1979). "Inversion of Subject NP and AUX in Children's Questions", Journal of Psycholinguistic Research 8: 333-341.

Iwakura, K. (1977). "The Auxiliary System in English", Linguistic Analysis 3: 101-136.

Jackendoff, R. (1972). Semantic Interpretation in Generative Grammar, MIT Press, Cambridge, MA.

Jackendoff, R. (1977). X'-Syntax: A Study of Phrase Structure, MIT Press, Cambridge, MA.

Jackendoff, R. (1983). Semantics and Cognition, MIT Press, Cambridge, MA.

Jacobson, P. & G. Pullum, eds. (1982). The Nature of Syntactic Representation, Reidel, Dordrecht.

Jaeggli, O. (1980). "Remarks on to-Contraction", Linguistic Inquiry 11: 239-245.

Jaeggli, O. (1982). Topics in Romance Syntax, Foris publications, Dordrecht.

Jaeggli, O. (1984). "Subject Extraction and the Null-Subject Parameter", in Proceedings of the Fourteenth Annual Meeting of the North Eastern Linguistics Society, University of Massachusetts, Amherst, MA.

Johnson, C. (1981). Children's Questions and the Discovery of Interrogative Syntax, unpublished Stanford University PhD dissertation.

Kaisse, E. (1985). Connected Speech: The Interaction of Syntax and Phonology, Academic Press, New York.

Kaper, W. (1984). Child Language: A Language Which Does Not Exist ?, Foris publications, Dordrecht.

Kaplan, R. & J. Bresnan (1982). "Lexical Functional Grammar: a Formal System for Grammatical Representation", in J. Bresnan, ed.

Kayne, R. (1972). "Subject Inversion in French Interrogatives", in J. Casagrande & B. Saciuk, eds., Generative Studies in Romance Languages, Newbury House, Rowley, MA.

Kayne, R. (1975). French Syntax: the Transformationakl Cycle, MIT Press, Cambridge, MA.

Kayne, R. (1984). Connectednes and Binary Branching, Foris publications, Dordrecht.

Kayne, R. & J.-Y. Pollock (1978). "Stylistic Inversion, Successive Cyclicity and Move NP in French", Linguistic Inquiry 9: 595-621.

Kiparsky, P. (1982). "From Cyclic Phonology to Lexical Phonology" in H. van der Hulst & N. Smith, eds., The Structure of Phonological Representations (Part I), Foris publications, Dordrecht.

Klein, S. (1982). Syntactic Theory and the Developing Grammar: Reestablishing the Relationship between Linguistic Theory and Data from Language Acquisition, unpublished UCLA PhD dissertation.

Klima, E. (1964). "Negation in English", in J. Fodor & J. Katz, eds., The Structure of Language, Prentice-Hall, Englewood Cliffs, New Jersey.

Klima, E. & U. Bellugi (1966). "Syntactic Regularities in the Speech of Children" in J. Lyons & R. Wales, eds., Psycholinguistic Papers, Edinburgh University Press, Edinburgh.

Koopman, H. (1983). "Control from COMP and Comparative Syntax", The Linguistic Review 2: 365-385.

Koopman, H. (1984) The Syntax of Verbs, Foris publications, Dordrecht.

Koster, J. & R. May (1982). "On the Constituency of Infinitives", Language 58: 116-143.

Kuczaj, S. (1976). "Arguments Against Hurford's AUX-Copying Rule", Journal of Child Language 3: 423-427.

Kuczaj, S. (1978). "Why Do Children Fail to Overgeneralize the Progressive Inflection ?", Journal of Child Language 5: 167-171.

Kuczaj, S. & N. Brannick (1979). "Children's Use of the WH-Question Modal Auxiliary Placement Rule", Journal of Experimental Child Psychology 28: 43-67.

Kuczaj, S. & M. Maratsos (1975). "What Children Can Say Before They Will", Merril-Palmer Quarterly 21: 89-111.

Kuczaj, S. & M. Maratsos (1983). "The Initial Verbs of Yes-No Questions: a Different Kind of General Grammatical Category", Developmental Psychology 19: 440-444.

Labov, W. (1972). Sociolinguistic Patterns, University of Pennsylvania Press, Philadelphia.

Labov, W., P. Cohen, C. Robins, & J. Lewis (1968). A Study of the Non-Standard English of Negro and Puerto-Rican Speakers in New York City, Cooperative Research Report # 3288, U.S. Dept. of Health & Welfare.

Labov, W. & T. Labov (1978). "Learning the Syntax of Questions", in R. Campbell & P. Smith, eds., Recent Advances in the Psychology of Language: Formal and Experimental Approaches, Plenum Press, New York.

Ladusaw, W. (1981)."On the Notion Affective in the Analysis of Negative-Polarity Items", Journal of Linguistic Research 1: 1-16.

Lapointe, S. (1980a.). "A Lexical Analysis of the English Auxiliary Verb System", in T. Hoekstra, H. van der Hulst & M. Moortgat, eds., Lexical Grammar, Foris publications, Dordrecht.

Lapointe, S. (1980b.) "A Note on Akmajian, Steele and Wasow's Treatment of Certain Verb Complkement Types", Linguistic Inquiry 11: 770-787.

Lasnik, H. & M. Saito (1984). "On the Nature of Proper Government", Linguistic Inquiry 15: 235-289.

Lebeaux, D. (1987). "Comments on Hyams", in T. Roeper & E. Williams, eds.

Leijong, X. (1986). "Free Empty Category", Linguistic Inquiry 17: 75-93.

Levelt, W. (1979). "On Learnability: a Reply to Lasnik and Chomsky", ms., Nijmegen.

Levin, B. (1983). On the Nature of Ergativity, unpublished MIT PhD dissertation.

Levin, J. (1983). "Government Relations and the Structure of INFL", MIT Working Papers in Linguistics 5: 121-150.

Levin, J. and D. Massam (1985). "Surface Ergativity: Case/Theta Relations Reexamined", Proceedings of the Fifteenth Meeting of the North Eastern Linguistics Society, University of Massachusetts, Amherst, MA.

Li, C., ed. (1976). Subject and Topic, Academic Press, New York.

Li, C. & S. Thompson (1976). "Subject and Topic: a New Typology of Language", in C. Li, ed.

Lightfoot, D. (1979). Principles of Diachronic Syntax, Cambridge University Press, Cambridge.

Lobeck, A. (1986). Syntactic Constraints on VP Ellipsis, unpublished University of Washington PhD dissertation.

Lowenstamm, J. (1977). "Relative Clauses in Yiddish: a Case for Movement", Linguistic Analysis 3: 197-216.

MacNamara, J. (1982). Names for Things: a Study of Child Language, Bradford Books/MIT Press, Cambridge, MA.

MacWhinney, B. (1982). "Basic Processes in Syntactic Acquisition", in S. Kuczaj, ed., Language Development, Vol. 1: Syntax and Semantics, Erlbaum, Hillsdale, New Jersey.

Major, D. (1974). The Acquisition of Modal Auxiliaries in the Language of Children, Mouton, The Hague.

Maling, J. (1980). "Inversion in Embedded Clauses in Icelandic", Islenskt Mal 2: 175-193.

Maling, J. & A. Zaenen (1982). "A Phrase Structure Account of Scandinavian Extraction Phenomena" in P. Jacobson & G. Pullum, eds.

Manzini, M. R. & K. Wexler (1984). "Parameters, Learnability,

and Binding Theory", ms., University of California at Irvine.

Marantz, A. (1984). On the Nature of Grammatical Relations, MIT Press, Cambridge, MA.

Maratsos, M. (1974). "Children Who Get Worse at Understanding the Passive: a Replication of Bever", Journal of Psycholinguistic Research 3: 65-74.

Maratsos, M. (1978). "New Models in Linguistics and Language Acquisition" in M. Halle, J. Bresnan, & G. Miller, eds.

Maratsos, M. & S. Kuczaj (1978). "Against the Transformationalist Account: a Simpler Analysis of Auxiliary Overmarkings", Journal of Child Language 5: 337-345.

Maratsos, M. & M. Chalkley (1980). "The Internal Language of Children's Syntax: the Ontogenesis and Representation of Syntactic Categories", in K. Nelson, ed., Children's Language, Vol. 2, Gardner, New York.

Marcus, M. (1980). A Theory of Syntactic Recognition for Natural Language, MIT Press, Cambridge, MA.

Massam, D. (1985). Case Theory and the Projection Principle, unpublished MIT PhD dissertation.

May, R. (1977). The Grammar of Quantification, Indiana University Linguistics Club, Bloomington, Indiana.

May, R. (1985). Logical Form, MIT Press, Cambridge, MA.

Mayer, , J., A. Erreich & V. Valian (1978). "Transformations, Basic Operations, and Language Acquisition", Cognition 6: 1-13.

McCawley, J. (1971). "Tense and Time Reference in English", in C. J. Fillmore & D. T. Langendoen, eds., Studies in Linguistic Semantics, Holt, Winston & Rinehart, New York.

McCawley, J. (1975). "The Category Status of English Modals", Foundations of Language 12: 597-601.

McClelland, D., Rumelhart, J. (1981). "An Interactive Activation Model of Perception, Part I.", Psychological Review 88: 375-407.

McClelland, D., Rumelhart, J. & the PDP Research Group (1986). Parallel Distributed Processing, Volume 2: Psychological and Biological Models, Bradford Books/MIT Press, Cambridge, MA.

McCloskey, J. (1979). Transformational Syntax and Model-Theoretic Semantics, Reidel, Dordrecht.

McCloskey, J. (1983). "A VP in a VSO Language ?", in G. Gazdar, E. Klein & G. Pullum, eds., Order, Concord, and Constituency, Foris publications, Dordrecht.

McCloskey, J. & K. Hale (1984). "On the Syntax of Person-Number Inflection in Modern Irish", Natural Language and Linguistic Theory 1: 487-533.

Menyuk, P. (1969). Sentences Children Use, MIT Press, Cambridge, MA.

Moerk, E. (1980). "Relations betwen Parental Input Frequencies and Children's Language Acquisition: a Reanalysis of Brown's Data", Journal of Child Language 7: 105-118.

Moerk, E. (1981). "To Attend or Not to Attend to Unwelcome Reanalyses ? A Reply to Pinker", Journal of Child Language 8: 627-631.

Miller, W. and S. Ervin (1964). "The Development of Grammar in Child Language", in U. Bellugi & R. Brown, eds.

Morin, Y.-C. (1979). "Some Myths about Pronominal Clitics in French", Linguistic Analysis 8: 95-109.

Nash, D. (1980). Topics in Warlpiri Grammar, unpublished MIT PhD dissertation.

Nelson, K. (1973). Structure and Strategy in Learning to Talk, Monographs of the Society for Research in Child Development 38.

Newport, E., H. Gleitman & L. Gleitman (1977). "Mother, I'd Rather Do It Myself: Some Effects and Noneffects of Maternal Speech Style", in C. Snow and C. Ferguson, eds., Talking to Children: Language Input and Acquisition, Cambridge University Press, Cambridge.

Ninio, A. & J. Bruner (1978). "The Achievement and Antecedents of Labelling", Journal of Child Language 7: 565-573.

Nishigauchi, T. & T. Roeper (1987). "Deductive Parameters" in T. Roeper & E. Williams, eds.

Obenauer, H.-G. (1984). "Connectedness, Variables, and Stylistic Inversion in French", in J. Gueron, H.-G. Obenauer, & J.-Y. Pollock, eds., Grammatical Representation, Foris publications, Dordrecht.

O'Grady, W. (1987). Principles of Grammar and Learning, University of Chicago Press, Chicago.

Park, T.-Z. (1971). "The Acquisition of German Syntax", ms., University of Berne.

Pesetsky, D. (1981). "Complementizer-Trace Phenomena and the Nominative Island Condition", The Linguistic Review 1: 297-343.

Pesetsky, D. (1982) Paths and Categories, unpublished MIT PhD dissertation.

Peters, A. (1983). The Units of Language Acquisition, Cambridge University Press, Cambridge.

Phinney, M. (1981). Syntactic Constraints and the Acquisition of Embedded Sentential Complements, unpublished U.Mass. PhD dissertation.

Pinker, S. (1979). "Formal Models of Language Learning", Cognition 7: 217-283.

Pinker, S. (1981). "On the Acquisition of Grammatical Morphemes", Journal of Child Language 8: 477-484.

Pinker, S. (1984). Language Learnability and Language Development, Harvard University Press, Cambridge, MA.

Pinker, S. & D. Lebeaux (1981). "A Learnability-Theoretic Approach to Children's Language", ms., Harvard University.

Platzack, C. (1983). "Germanic Word-Order and the COMP/INFL Parameter", in Working Papers in Scandinavian Syntax 2, University of Trondheim.

Platzack, C. (1986). "The Position of the Finite Verb in Swedish", in H. Haider & M. Prinzhorn, eds.

Pollock, J.-Y. (1987). "On Verb Raising", paper given at the 1987 GLOW Conference, Venice.

Postal, P. (1966). "On So-called Pronouns in English", in F. Dineen, ed., Nineteenth Monograph on Languages and Linguistics, Georgetown University Press, Washington, D.C.

Postal, P. and G. Pullum (1982). "The Contraction Debate", Linguistic Inquiry 13: 122-138.

Prideaux, G. (1976). "A Functional Analysis of English Question Acquisition: a Response to Hurford", Journal of Child Language 3: 417-422.

Pullum, G. (1985). "Assumimg Some Version of X-Bar Theory", in Papers from the General Session of the Twenty-First

Regional Meeting of The Chicago Linguistics Society, University of Chicago.

Pullum. G. & D. Wilson (1977). "Autonomous Syntax and the Analysis of Auxiliaries", Language 53: 435-463.

Pye, C. (1983). "Mayan Telegrapheses: Intonational Determinants of Inflectional Development in Quiche Mayan", Language 59: 583-604.

Reuland, E. (1983). "Governing -ing", Linguistic Inquiry 14: 101-136.

Riemsdijk, H. van (1978). A Case Study in Syntactic Markedness, Foris publications, Dordrecht.

Riemsdijk, H. van & E. Williams (1981). "NP-Structure", The Linguistic Review 1: 171-217.

Rivero, M.-L. (1978). "Topicalization and WH-Movement in Spanish", Linguistic Inquiry 9: 513-517.

Rivero, M.-L. (1980). "On Left-Dislocation and Topicalization in Spanish", Linguistic Inquiry 11: 363-393.

Rizzi, L. (1982). Issues in Italian Syntax, Foris publications, Dordrecht.

Rizzi, L. (1984). "On the Status of Subject Clitics in Romance", in O. Jaeggli & C. Silva-Corvalan, eds., Proceedings of the Fourteenth Linguistic Symposium on Romance Languages, Foris publications, Dordrecht.

Rizzi, L. (1986). "Null Objects in Italian and the Theory of pro", Linguistic Inquiry 17: 501-557.

Roberge, Y. (1986). The Syntactic Recoverability of Null Arguments, unpublished University of B.C. PhD dissertation.

Roberts, I. (1985). "Agreement Parameters and the Development of English Modal Auxiliaries", Natural Language and Linguistic Theory 3: 21-58.

Rochemont, M. (1986). Focus in Generative Grammar, John Benjamins, Amsterdam.

Roeper, T. & E. Williams, eds. (1987). Parameter Setting, Reidel, Dordrecht.

Rosenbaum, P. (1967). The Grammar of English Predicate Complement Constructions, MIT Press, Cambridge, MA.

Ross, J. R. (1967). Constraints on Variables in Syntax,

unpublished MIT PhD dissertation.

Ross, J. R. (1969). "Auxiliaries as Main Verbs", in W. Todd, ed., Studies in Philosophical Linguistics, Evanston, Illinois.

Ross, J. R. (1972). "Double-ing", Linguistic Inquiry 3: 61-86.

Rothstein, S. (1983). The Syntactic Form of Predication, unpublished MIT PhD dissertation.

Safir, K. (1982). "Inflection-Government and Inversion", The Linguistic Review 417-467.

Safir, K. (1985). Syntactic Chains, Cambridge University Press, Cambridge.

Safir, K. (1986). "Relative Clauses in a Theory of Binding and Levels", Linguistic Inquiry 17: 663-689.

Safir, K. (1987). "Comments on Wexler and Manzini", in T. Roeper & E. Williams, eds.

Safir, K. & D. Pesetsky (1981). "Inflection, Inversion, and Subject Clitics", in Proceedings of the Eleventh Meeting of the North Eastern Linguistics Society, University of Massachussets, Amherst, MA.

Sag, I. (1980). Deletion and Logical Form, Garland, New York.

Sag, I. (1982). The Syntax and Semantics of Verb-Phrase Ellipsis, Elsevier, New York.

Saito, M. (1985). Some Asymmetries in Japanese and their Theoretical Implications. unpublished MIT PhD dissertation.

Schachter, P. (1980). Explaining Auxiliary Order, Indiana University Linguistics Club, Bloomington, Indiana.

Schlesinger, J. M. (1971). "Learning Grammar; from Pivot to Realization Rule", in R. Huxley & E. Ingram, eds.

Schmerling, S. (1977). "The Syntax of English Imperatives", ms., University of Texas at Austin, Texas.

Sheldon, A. (1974). "The Role of Parallel Function in the Acquisition of Relative Clauses in English", Journal of Verbal Learning and Verbal Behaviour 13: 272-281.

Shieber, S. (1985). "Evidence Against the Context-Freeness of Natural Language", Linguistics and Philosophy 8: 333-343.

Simpson, J. (1982). Control and Predication in Warlpiri,

unpublished MIT PhD dissertation.

Slobin, D. (1985). The Cross-Linguistic Study of Language Acquisition, Erlbaum, Hillsdale, New Jersey.

Smith, C. S. (1970)."An Experimental Approach to Linguistic Competence", in J. R. Hayes, ed.

Smith, C. S. & A. van Kleek (1986). "Linguistic Complexity and Performance", Journal of Child Language 13: 389-408.

Snow, C. (1986). "Conversations with Children", in P. Fletcher & M. Garman, eds.

Speas, M. (1984). "Complement Selection and Inversion", ms., MIT.

Speas, M. (1986). Adjunctions and Projections in Syntax, unpublished MIT PhD dissertation.

Sportiche, D. (1983). Structural Invariance and Symmetry in Syntax, unpublished MIT PhD dissertation.

Sproat, R. (1983). "INFL and the Configurationality of VSO Languages", in Proceedings of the Fourteenth Meeting of the North Eastern Linguistics Society, University of Massachusetts, Amherst, MA.

Steele, S., A. Akmajian, R. Demers, E. Jelinek, R. Oehrle & T. Wasow (1981). An Encyclopedia of AUX: a Study in Cross-Linguistic Equivalence, MIT Press, Cambridge, MA.

Stemberger, J. (1982a.) "The Nature of Segments in the Lexicon: Evidence from Speech Errors", Lingua 46: 43-67.

Stemberger, J. (1982b.) "Syntactic Errors in Speech", Journal of Psycholinguistic Research 11: 313-339.

Stephany, U. (1986). "Modality", in P. Fletcher & M. Garman, eds.

Sternefeld, W. (1985). "On Case and Binding Theory", in J. Toman, ed.

Stowell, T. (1981). Origins of Phrase-Structure, unpublished MIT PhD dissertation.

Stowell, T. (1983). "Subjects Across Categories", The Linguistic Review 2: 285-312.

Stowell, T. (1987). "X-bar Theoretic Semantics and the DP Hypothesis", paper given at the 1987 GLOW conference, Venice.

Stromswold, K. & S. Pinker (1986). "Questions about WH-Questions", paper given at the Eleventh Annual Conference on Language Development, University of Boston, Boston, MA.

Tajima, K. (1987). "WH-Q/WH-Rel Asymmetries and Conditions on Chains", ms., University of Washington.

Tappe, H.-T. (1982). "VP and Coherent Infinitives in German", ms., University of Gotingen.

Taraldsen, K. (1981). "The Theoretical Interpretation of a Class of Marked Extractions", in A. Beletti, L. Brandi & L. Rizzi, eds., A Theory of Markedness in Generative Grammar, Scuola Normale Superiore, Pisa.

Tavakolian, S., ed. (1981). Language Acquisition and Linguistic Theory, MIT Press, Cambridge, MA.

Thrainsson, H. (1986). "V1, V2, V3 in Icelandic", in H. Haider & M. Prinzhorn, eds.

Toman, J., ed. (1985). Studies in German Grammar, Foris publications, Dordrecht.

Torrego, E. (1984). "On Inversion in Spanish and Some of its Effects", Linguistic Inquiry 15: 103-129.

Travis, L. (1984). Parameters and Effects on Word-Order Variation, unpublished MIT PhD dissertation.

Ultan, R. (1978). "Some General Characteristics of Interrogative Systems", in J. Greenberg, C. Ferguson, & E. Moravcsik, eds., Universals of Human Language,

Walli, E. (1984). "Scope Assignment and Negation", ms., MIT.

Webelhuth, G. (1986b.). "Some Data on the Verb-Object Relation in German", Linguistic Inquiry 17: 772-776.

Weerman, F. (1983). "Bouman revisited: V2 Verschijnselen in een Moderne Historische Taalkunde", ms.

Weinberg, A. (1987). "Comments on Borer and Wexler" in T. Roeper & E. Williams, eds.

Weist, R. (1986). "Tense and Aspect", in P. Fletcher & M. Garman, eds.

Wells, G. (1979). "Learning and Using the Auxiliary Verb in English" in V. Lee, ed., Language Development: a Reader, Croon Helm, London.

Wells, G. (1985). Language Development in the Pre-School Years,

Cambridge University Press, Cambridge.

Wells, G. (1986)."Variation in Child Language", in P. Fletcher & M. Garman, eds.

Wexler, K. & P. Culicover (1980). Formal Principles of Language Acquisition, MIT Press, Cambridge, MA.

Wexler, K. & M. R. Manzini (1987). "Parameters and Learnability in Binding Theory", in T. Roeper & E. Williams, eds.

White, L. (1982). Grammatical Theory and Language Acquisition, Foris publications, Dordrecht.

Williams, E. (1977). "Discourse and Logical Form", Linguistic Inquiry 8: 101-139.

Williams, E. (1978). "Across-the-Board Rule Application", Linguistic Inquiry 9: 31-43.

Williams, E. (1980). "Predication", Linguistic Inquiry 11: 203-238.

Williams, E. (1981). "Language Acquisition, Markedness, and Phrase-Structure" in S. Tavakolian, ed.

Williams, E. (1982). "The NP Cycle", Linguistic Inquiry 13: 277-295.

Williams, E. (1983). "Against Small Clauses", Linguistic Inquiry 14: 287-308.

Williams, E. (1987). "Introduction", in T. Roeper & E. Williams, eds.

Wittgenstein, L. (1953). Philosophical Investigations, Basil Blackwell, Oxford.

Zagona, K. (1982). Government and Proper Government of Verbal Projections, unpublished University of Washington PhD dissertation.

Zubizaretta, M.-L. (1982). On the Relationship of the Lexicon to Syntax, unpublished MIT PhD dissertation.

Zwicky, A. & G. Pullum (1983). "Cliticization vs. Inflection: English n't", Language 59: 502-513.